New trends in integrated science teaching

Volume IV

Evaluation of integrated science education

Integrated science

The teaching of basic sciences

unesco

The teaching of basic sciences

Titles in this series

A survey of the teaching of physics at universities. 1966

New trends in biology teaching.
Vol. I: 1967
Vol. II: 1969
Vol. III: 1971

New trends in chemistry teaching.
Vol. I: 1967
Vol. II: 1969
Vol. III: 1972
Vol. IV: 1975

New trends in mathematics teaching.
Vol. I: 1966
Vol. II: 1970
Vol. III: 1972

New trends in physics teaching.
Vol. I: 1968
Vol. II: 1972
Vol. III: 1976

Mathematics applied to physics. 1970

New trends in integrated science teaching.
Vol. I: 1971
Vol. II: 1973
Vol. III: 1974
Vol. IV: 1977

New trends in integrated science teaching:
evaluation of integrated science education

Volume IV

Edited by
David Cohen,
Associate Professor of Education,
Macquarie University,
New South Wales,
(Australia)

unesco

The opinions expressed in this publication are the responsibility of the authors and of the editor, and do not necessarily reflect those of Unesco. Furthermore, the designations employed and the presentation of the material in this work do not imply the expression of any opinion whatsoever on the part of Unesco concerning the legal status of any country or territory, or of its authorities, or concerning the delimitations of the frontiers of any country or territory.

Published in 1977 by the United Nations Educational, Scientific and Cultural Organization, 7 place de Fontenoy, 75700 Paris
Printed by Floch, Mayenne

ISBN 92-3-101440-4
Spanish edition : 92-3-301440-1

Preface

This publication is the fourth work on the subject of integrated science education in the Unesco series The teaching of basic sciences. The first two volumes in the series (New Trends in Integrated Science Teaching, Vol. I and II) attempted to exemplify and to analyse world-wide trends in this rapidly-evolving field. The third volume addressed itself to the crucial problem of the education of teachers for integrated science. It was based on the proceedings of a conference on this subject organized by the International Council of Science Unions (ICSU) Committee on the Teaching of Science, with Unesco support, at the University of Maryland (United States) in April 1973.

This present volume attempts to tackle another area vital to the development of integrated science education, namely that of evaluation. This area was singled out by the Maryland Conference as one where little information was available and which required urgent attention. Unesco has also received many requests for reports of on-going evaluation projects in this field and for information on the methodology of testing and examining in integrated science. It has, in addition, been asked to assist in designing evaluation programmes, usually in situations where there are severe limitations on funds and personnel.

Accordingly, working in collaboration with the International Council of Associations for Science Education (ICASE), Unesco planned a two-part programme activity, consisting of a Symposium on "Evaluation of Integrated Science Education" and a publication based on the proceedings of this Symposium. A consultant, Dr. D. Cohen of Macquarie University (Australia), was invited to assist in the preparation of the Symposium, including the selection of participating specialists, and to edit the final publication.

The Symposium was held at Oxford (United Kingdom) for a four-day period in December 1975, concurrently with the General Assembly of ICASE. A series of fourteen manuscripts had been commissioned for the Symposium and their authors first met together to discuss and comment on each others' contributions. They then met with the ICASE delegates to discuss the relevance and applicability of their work to the problems and conditions of integrated science teaching in the thirty-eight countries represented. Following this, the authors met together again to attempt to incorporate the feedback received and to bring the manuscripts together to form a unified whole. While complete unanimity was found to be impossible, a very considerable measure of agreement was achieved. As a result of the feedback, an additional chapter on "Evaluating Students' Progress in Integrated Science - Public Examinations" was commissioned.

The publication seeks to present a thorough review and exemplification of current knowledge and practice in the field, and the authors have attempted a wide coverage of material and ideas. The first part of the book consists of chapters which review and analyse distinct aspects of the topic, as revealed by the literature throughout the world and the accounts of experimental projects and international, regional and national conferences. The second part comprises seven case studies of on-going evaluation projects in different parts of the world. These include accounts of the problems encountered when attempting to evaluate integrated science education in a concrete situation and some suggestions for tackling these problems.

The readership for which this volume is intended is a wide one. It is hoped that it will be found useful by practising teachers and teachers-in-training as well as by teacher educators, science curriculum development and research workers, and Ministry of Education officials. It is suggested that selected portions of the book be used as resource material in science education workshops and conferences organized by professional bodies concerned with science education.

Appreciation is expressed to the International Council of Association for Science Education and especially to its secretary, Mr. D. Chisman, who undertook the organizational work for the Oxford Symposium. Grateful thanks are due to all the authors of commissioned manuscripts and, in particular, to Professor J.F. Kerr who responded to the request to write a manuscript at very short notice. Particular acknowledgement is made of the devoted work of the Editor. Dr. D. Cohen. Tribute is also due to Mr. R. W. Morris of the University of Sussex and his colleagues Mrs. Brenda Harmer and Mrs. Brenda Haw who worked together as a team to convert the manuscript into camera-ready copy for the printer.

The views expressed are the responsibility of the authors, and do not necessarily reflect the views of Unesco.

Contents

1 Evaluation in integrated science teaching — an introduction

David Cohen
Macquarie University, Australia

SUMMARY

The growing interest in integrated science teaching has sharpened the need to clarify the meanings and interpretations of integration and accelerated the development of new strategies of evaluation in integrated science teaching. This introductory chapter briefly reviews some alternatives and indicates the nature of the chapters to follow.

INTRODUCTION

The purpose of this publication is to provide a state-of-the-art report concerning current thinking and new trends in evaluation related to integrated science teaching. With the renewed rethinking about the nature of integrated science teaching and a renewed interest in reviewing alternatives concerning the nature and roles of evaluation, the marriage of evaluation with integrated science teaching is very much in the honeymoon stage. Consequently, in many respects, discussions remain necessarily pioneering if not speculative. The current evolutionary stage represents one good reason for producing a status document just three-quarters of the way through the twentieth century. Are there any laws, panaceas, recipes or generalizations emerging about evaluation in integrated science teaching?

Unlike the exact sciences, education can boast no National Standards Laboratories. Universal agreements and enactments about education and its terminology do not exist. Even the number of definitions for a particular educational expression continue to proliferate and each definition may be subjected to widely differing interpretations. The range of alternative interpretations thus generated expands exponentially. A first series of decision-making processes is concerned with a need to develop or choose definitions.

In considering Evaluation in Integrated Science Teaching, a detailed review of the terms "evaluation" and "integrated science teaching" could fill the whole volume. How might we define each? How may we interpret each? What varying influences do varying international contexts have upon them? What is the effect of interrelating the terms? What alternative strategies for evaluation are available? How might they be implemented? The resultant semantic, philosophic and educational analyses produce numerous options. The clarification of the concepts of integrated science teaching and of evaluation is akin to the elusive mirage, for as we stride forward, the obvious answers recede, only to reveal new and broader issues and options. Beneath the surface of an initial apparent clarity of the terms lie complex networks which make attempted generalizations about "evaluation in integrated science teaching" highly speculative.

[1] Acknowledgements are made to Ms Ruth Cohen, Dr. Barry Fraser, and Dr. Robert Precians for their helpful criticisms.

Integration

Consider first the spectrum of acceptable alternative meanings of "integration" as the term relates to science teaching. "Integration" is generally applied to the relationships within and between subjects more often taught separately, so that it refers to a horizontal relationship within a curriculum at a particular educational year level. A key intent of those supporting the concept of integration is to provide a unified view. An alternative view of integration is to consider it as a recombination of subjects or other units into which the whole curriculum was earlier divided in order to reflect, in part, the existence of subject disciplines as areas of study, research and activity.

Following an extensive review of the literature, d'Arbon summarized in this way:

> " 'Integration', when applied to science courses, means that the course is devised and presented in such a way that students gain the concept of the fundamental unity of science; the commonality of approach to problems of a scientific nature; and are helped to gain an understanding of the role and function of science in their everyday life, and the world in which they live.
>
> An integrated science course eliminates the repetition of subject matter from the various sciences and does not recognize the traditional subject boundaries when presenting topics or themes.
>
> Integrating principles are intended to produce a course which
> (i) is relevant to student needs and experiences;
> (ii) stresses the fundamental unity of science;
> (iii) lays adequate foundations for subsequent specialist study; and
> (iv) adds a cultural dimension to science education." [1]

This interpretation reflects the view that the several sciences are characterized by a common methodology and that their contents represent a whole and become more meaningful when interrelated. This implies that the subdivision of the whole curriculum into separate subjects may be considered as "disintegration", as may also the subdivision of science into the separate sciences. Integration may then be regarded as the removal of boundaries between subject subdivisions or categories. The extent of integration might vary from the re-combination of two subjects into a single subject (e.g. aspects of physics and chemistry into physical chemistry) to the re-combination of several subjects into a single subject (e.g. biology, chemistry, earth sciences and physics integrated to form "integrated" science). More widely again, by integrating also some concepts or materials from the non-sciences, "integrated studies" would result.

However, there is also a strong school of thought favouring the retention of separate science subjects to reflect the "disciplines", based partly upon the view that people need conceptual structures to avoid a confusion of facts. Such structures are said to be based upon the integrity both of content and of the specific approaches to inquiry unique to each discipline. Proponents of the disciplines perceive integration as an undesirable trend, if not a vain hope. They argue that the separate study of chemistry, physics, biology and earth sciences as scientific disciplines is necessary to understand the content and methodology of each, and that attempts at integration represent an erosion of intellectual integrity. It could be further argued that the separate sciences would lose their own individual identities and so become unrecognizable within "integrated science" and that complete fusion (e.g. to create themes) would make it difficult to identify and study adequately the separate disciplines within integrated science teaching.

In summary, consensus is lacking concerning the "pros" and "cons" of the study of science as comprising "separate subjects", or as "integrated science". Criteria to assist decision-makers should include the results of careful appraisals of the context, over-all curriculum objectives, functions and availability both of appropriately educated teachers and of physical resources.

[1] J.A. d'Arbon, "A Study of the Concept of Integration of Science Subjects in Secondary Schools". Unpublished thesis presented as partial requirement for the degree of Master of Science, Macquarie University, North Ryde, N.S.W., 1972.

Evaluation

The range of platforms related to "integration" is parallelled in number and diversity by the range of views related to evaluation. The growing awareness of the complexities of evaluation were highlighted by discussions at the Paris Symposium during the preparation of New Trends in Integrated Science Teaching : Volume II[1] (Richmond, 1973) and again during the conference at Maryland on "The Education of Teachers for Integrated Science - Teaching Science for Today's Science" related to New Trends in Integrated Science Teaching : Education of Teachers[2] (Richmond, 1974). These led to convene the 1975 Oxford Symposium, devoted entirely to a study of evaluation applied to integrated science teaching.

The study of evaluation in general has been the focus of a rapid growth of international interest, mirrored worldwide in the meteoric rise in the number of papers published and the number of conferences held concerning evaluation. An index to this growth is the proliferation of publications concerning various aspects of evaluation, especially in the last decade. As examples, there have appeared numbers of anthologies, a series of AERA Monographs, a sourcebook, an encyclopedia, a yearbook, a handbook, a Commission report, a review of research, a treatise, a host of books, and scores of journal articles, including whole theme issues (see bibliography for details). The reasons for this growth of interest and activity are diverse. They include increased funding for education with the related pressures for accountability for the expenditure of these funds, a growing disenchantment with the lack of pay-off from traditional methods of evaluation, the generation of a new range of strategies for attacking the problems of evaluation, increased availability of professionals with competencies to handle evaluation studies and an emphasis on the improvement of educational quality and verification of this improvement through the provision of evidence.

The expansion of interest in and views about evaluation have led to polarizations of views on a number of dimensions describing the nature, scope and strategies of evaluation. Table 1 represents the range of views along eight different dimensions.

[1] Peter E. Richmond (ed), New Trends in Integrated Science Teaching : Volume II, Unesco, Paris, 1973.

[2] Peter E. Richmond (ed), New Trends in Integrated Science Teaching : Education of Teachers, Volume III, Unesco, Paris, 1974.

Table 1. Some dimensions in evaluation.

Dimensions	Range of views represented along continua	
Data provision to decision-makers	Data are pre-selected by evaluators to provide to decision-makers	Broad, diverse arrays of data are provided, from which decision-makers make their own choices.
Nature of data	Normative data; value-based (e.g. examination scores)	Descriptive data; value-free (e.g. classroom anecdotal records).
Referents for evaluation of student progress	Numerical or letter grades used to compare student performances (norm-referenced)	Student progress related to objectives (criterion-referenced).
Bases of evaluation	Heavy emphasis upon instruments to provide standardized test scores	Greater reliance upon teachers who exercise spontaneous subjective judgements.
Specificity of aspects evaluated	Aspects to be evaluated fully specified (often behaviourally) and pre-determined (means-end evaluation)	Emerging developments result in progressive modification of evaluation strategies (including evaluation of unintended outcomes).
Breadth of evaluation	Evaluation limited to assessment of student achievement	Evaluation embraces assessment of students, teachers, curricula, learning processes, materials, teaching skills, etc.
Process/product evaluation	Evaluation concerned with ends only (i.e. products)	Evaluation concerned also with processes of development, implementation and evaluation.
Phasing of evaluation	Formative and summative evaluation considered as two-phase, separate and comprehensive scheme	Evaluation seen as continuous, overlapping and interacting processes, including reflective evaluation as well as formative and summative evaluation.

Each dimension in Table 1 represents a continuum of views. For example, the first dimension relates to the provision of data to the decision-makers. The views along this continuum range from those who consider that evaluators make a pre-selection of which data to provide, to those evaluators who consider that a mass of diverse data should be provided from which the decision-makers choose.

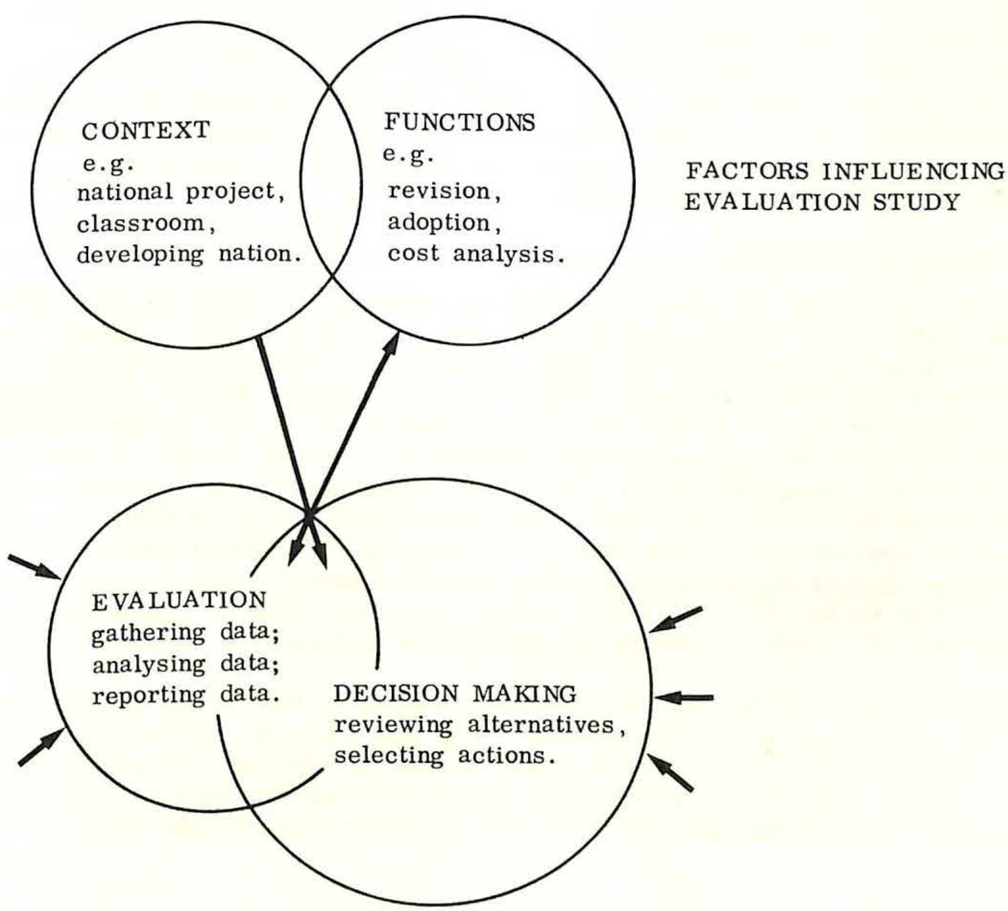
As used in everyday speech, evaluation is concerned with judgements about values, with questions of the nature: "How good is ...?" and "What is the worth of ...?". However, these has been a growing tendency to separate the judgemental from the information-provision aspects of evaluation. Thus, evaluation has been defined in this way:

"Educational evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives".[1]

This view of evaluation implies that it is the prime tool for decision-makers who include the students, teachers, parents and community, education departments, governmental agencies, project teams and publishers. Not only is evaluation concerned with gathering, analysing and reporting data in order to improve decision-making, but evaluators may also help decision-makers to articulate their value positions, without the evaluators themselves becoming decision-makers, as Tawney contends in Chapter 2.

Evaluation may be considered as part of a complex network, interacting with the context and functions which the evaluation is to serve, as well as the decision-making processes (see Fig. 1). The potential functions to be served by the decisions to be made should help evaluators decide upon the most appropriate data to be gathered. Figure 1 shows diagrammatically the complex network of interactions between the concern of evaluators to service the decision-makers with data appropriate for their differing contexts and functions.

Fig. 1. Evaluation as a complex interacting network.



[1] Daniel L. Stufflebeam, et al. Educational Evaluation and Decision Making, F.E. Peacock Publishers Inc., Itasca, Ill., 1971.

The context of evaluation will heavily influence the strategies selected. For example, the methods of evaluation will differ according to context. Contexts will vary with the type of society (with its implicit values), the economic background of schools, the availability of adequate funding and personnel, the extent of centralization and the stage of development reached in the educational system (e.g. level of teachers).

Another kind of variation in context relates to the magnitude of the evaluation activity. In the case of a large-scale evaluation (as might be needed for a national, state or curriculum project activity), members of the development team and evaluators may each have differing perceptions about evaluation, and differing commitments to different approaches. Factors such as interpersonal relationships and perceived status of the opinion-holders may be one influence upon the acceptance or rejection of their views. Judgements may be periodic, formal (e.g. based on standardized or external testing) and "objective". An example of small-scale evaluation is the teacher in a classroom, making continuous, informal, spontaneous and subjective judgements.

Just as evaluation procedures are likely to differ in differing contexts, so too are they likely to vary according to the range of decisions to be based upon them. The functions of the decisions may be diverse, relating to such questions as whether to revise aspects of a curriculum in its developmental stages, whether to change approaches to implementation, or to adopt/adapt/reject some "package" approach to curriculum materials or teaching strategies. Criteria upon which such decisions might be based include costs, predicted educational benefits and feasibility. These decisions may vary also according to the context in which they are made.

The complexities of the interactions represented in Figure 1 make it clear that there can be no unique approach, no sets of rules or recipes or "pat" answers to problems of evaluation. Even when a set of strategies is selected, these themselves are likely to change as the evaluation activities progress. What this present volume seeks to do is to elucidate some of the alternative approaches to evaluation in integrated science teaching.

Evaluation in Integrated Science Teaching

In planning this volume, an attempt was made to produce an exhaustive and logically comprehensive review of evaluation related to the various aspects of integrated science teaching. This necessitated contributions which analysed the strengths and limitations of evaluation (Tawney) and examined how evaluation could be used to improve the quality of decisions (Welch). Evaluation may be applied to a variety of aspects of integrated science teaching. In this volume, successive chapters describe the evaluation of integrated science curriculum materials (Bloch), of teaching skills (Sutton) and of students through the use of assessment instruments (Mayer and Richmond). Subsequent chapters review the use of teacher-developed approaches to evaluation (Ramsey) and assessment by means of external examinations (Kerr), and these are followed by Harlen's account of how evaluation may be used as the basis for adapting integrated science teaching experiences to individual pupils. The six national case studies which follow were selected to provide diverse examples of influences of differences in contexts and functions of evaluation in integrated science teaching.

As most of the approaches, techniques and examples contained in this volume are relevant to other areas as well as to integrated science teaching, it is hoped that the reviews which it contains will prove useful to both science teachers and evaluators in widening their educational horizons. May the volume itself be evaluated through the improved opportunities it might open, at least through integrated science teaching, for individual learners throughout the world.

BIBLIOGRAPHY

Anderson, Scarvia B., Ball, Samuel,. Murphy, Richard T. and Associates, Encyclopedia of Educational Evaluation, Jossey-Bass Publishers, San Francisco, 1973.

Bloom, Benjamin S., Hastings, J. Thomas, Madaus, George F., Handbook on Formative and Summative Evaluation of Student Learning, McGraw Hill, New York, 1971.

Denny, T. (Issued Ed), Review of Educational Research. Educational Evaluation, 40, 2, 1970, 181-324.

Provus, M., Discrepancy Evaluation, McCutchan Publishing Corp., Berkeley, California, 1971.

Stufflebeam, Daniel L. et al, Educational Evaluation and Decision Making, F.E. Peacock Publishers Inc., Itasca, Illinois, 1971.

Tyler, Ralph W. (ed), Educational Evaluation : New Rules, New Means, University of Chicago Press, Chicago, Illinois, 1969.

Walberg, Herbert J. (ed), Evaluating Educational Performance, McCutchan Publishing Corp., Berkeley, California, 1974.

Weiss, Joel (ed), Curriculum Theory Network, Monograph Supplement: Curriculum Evaluation: Potentiality and Reality, The Ontario Institute for Studies in Education, Ontario, Canada, 1972.

AERA Monograph Series on Curriculum Evaluation, Rand McNally and Co., Chicago.

Du Bois, Philip H. and Mays, G. Douglas (eds), Research Strategies for Evaluating Training, 1970.

Gallagher, James J., Nuthall, Graham A., and Rosenshine, Barak, Classroom Observation, 1970.

Grobman, Hulda, Evaluation Activities of Curriculum Projects, 1968.

Kraft, Richard H.P. et al, Four Evaluation Examples : Anthropological Economic Narrative and Portrayal, 1974.

Lindvall, C.M., Cox, Richard C., The IPI Evaluation Program, 1970.

Popham, W. James, Eisner, Eliot W., Sullivan, Howard J., Tyler, Louise L., Instructional Objectives, 1969.

Tyler, Ralph, Gagne, Robert and Scriven, Michael, Perspectives of Curriculum Evaluation, 1967.

2 The evaluation of integrated science education — its strengths and limitations, its scope and methodology

D. A. Tawney
University of East Anglia, United Kingdom

SUMMARY

Evaluation is undertaken to improve education, to reassure and, regrettably, because it is currently fashionable. Its strengths come from separating information-providing and option-probing from decision-making roles; its limitations come from the over-expectation it so often produces. The areas about which information should be provided include context, aims, processes and effects. Its methodology needs to be wide-ranging and is likely in practice to lie between the descriptive and quantitative extremes described.

Why Evaluate?

Through the rhetoric of justification emerge three main answers to the question: Why evaluate? Firstly, it is hoped that evaluation will help to improve education, whether it is used formatively to help a curriculum project develop its materials or to help a teacher develop his own teaching or, summatively, guiding a decision about which curriculum to adopt. After a decade and a half of accelerated curriculum development, education, including science education, has not changed as much as many had hoped. Evaluation is part of an attempt to go about development more sensitively and critically; evaluation is for quality control.

Secondly, evaluation is used to defend and to reassure. Funding bodies need confirmation that money has been well-spent, governments that a curriculum is worth disseminating, head teachers that pupils are not becoming "innocent victims of the poor pilot scheme" (as described for Malaysia by Sim, chapter 15) and the curriculum developers themselves that they have achieved something. Such a use of evaluation is understandable and at times necessary, but it should be regarded with caution as there is a danger of devoting resources to a study which will have little or no effect on decisions.

Thirdly, evaluations are commissioned because at the moment evaluation is fashionable; no good curriculum project can afford to be without one! It is not of course impossible for a study commissioned for so superficial a reason to turn out to be very valuable but unless an evaluation is commissioned carefully, with a full understanding of its purpose and likely outcomes, results can be disappointingly irrelevant.

The question: Why evaluate? is misleading, for everyone evaluates anyway. The question is whether this natural and inevitable aspect of most purposeful activities is better formalized, recognized as a separate activity, made the responsibility of particular people. Does this formalization improve the decisions made? Evaluators have to confess that there is as yet little evidence that it does.

Against evaluation as a separate activity, it must be argued that it is expensive; a summative evaluation may cost 5 per cent of a project's budget, a combined formative and summative evaluation 10 per cent. To spend less is to risk wasting money; for example, from a Scottish experience Jeffrey points out the limitations of using cheap labour involving students

working for higher degrees. Another argument is that too much information hinders developers' creativity, particularly if their confidence is sapped by unfavourable reports. However, although evaluators must be aware of these dangers, the argument represents a short-term view.

Readers must decide for themselves whether evaluation should be a formalized, separate activity. The subsequent pages, particularly the case studies, are intended to help this decision.

What are its strengths?

From the definition in Chapter 1, it is clear that evaluation as a separate activity is to do with the provision of information, with the intention of helping those who have to make decisions to make them more rationally. By separating the information provision from the decision-making, it is hoped that the information may be made more reliable. Evaluators are less involved, less committed to the materials or processes being studied, less anxious about the consequences of decisions. They have a more detached perspective and different values, more time for gathering information; teachers will tell evaluators things which are different from those they will tell the developers and the reports of the evaluators will have greater credibility.

Of course, it is quite possible for the developers themselves to adopt evaluation roles, deliberately distancing themselves from their creative roles; instruments such as tests and observation schedules are a good way of achieving this. It was the policy of the Nuffield Advanced Biological Science Project to share evaluation roles out among the developers[1] but small projects will do this of necessity as must teachers trying to evaluate their own work. However, it is difficult to play two roles simultaneously and the individual teacher may well be advised to seek the help of a colleague (see Chapter 5).

The value of a separate evaluator becomes clear if, following Stufflebeam[2], it is accepted that an evaluator should be concerned not only with the provision of information about what is but also with helping decision-makers articulate their values and explore future options, what might be. Many would agree with Sim that brainstorming with the decision-makers is part of evaluation.

The role of evaluator has opposite tendencies to that of the decision-maker; it encourages divergence and so deters a decision-maker from converging too soon, and it seeks to publish information, not trusting it unless it has been subject to public scrutiny, while decision-makers often prefer secrecy.

The ultimate strength of evaluation depends on a belief that decisions are better for being based on the maximum of relevant information and for being made after alternatives have been properly explored. It is the evaluator's job not only to supply this information but also to supply it at the right time and in a readily comprehensible form.

What are its limitations?

If evaluation is regarded as a panacea for all educational ills, it is bound to produce disappointment. Evaluation too often raises expectations which are too high; the following section is an attempt to lower them.

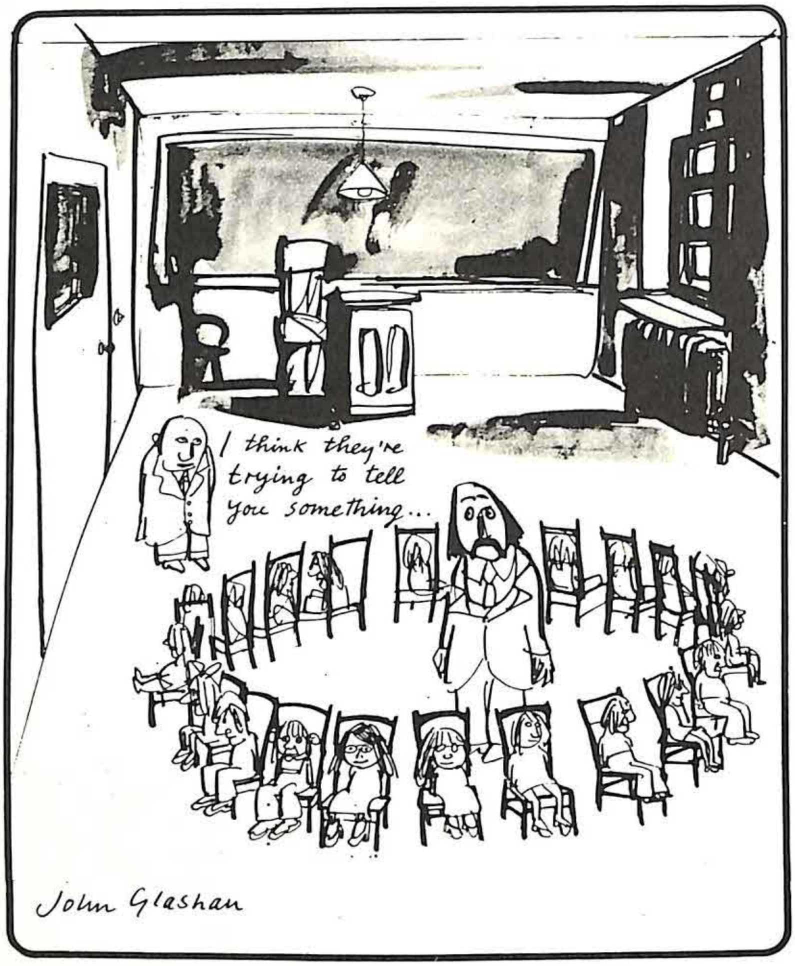
Evaluation in education cannot provide information with the reliability of the natural sciences.

This is contrary to current popular belief, and to that of evaluators in the past, who confidently adopted natural science research methodology. Evaluators are divided on whether the failure of past studies based on this methodology should lead to more sophisticated adaptations of it or to its abandonment. Even if numerical techniques were perfected, evaluation would still not be as decisive as some might hope because the aims of education depend on value judgements.

[1] P.J. Kelly, "Nuffield A-level Biological Science Project", in Evaluation in Curriculum Development: Twelve Case Studies (Schools Council Research Studies), Macmillan, London, 1973.

[2] Stufflebeam , et al, Educational Evaluation, op. cit.

The information provided by educational evaluation is unlikely to over-ride direct experience.



Teachers, administrators and developers build up direct experience by themselves observing and talking to participants. Reay's comment (Chapter 11) that there ... "is little evidence that the data were used by the decision-makers" ... is fairly universal; direct experience tends to win against a formal report. Evaluators need to interact with the decision-makers, "brainstorming" and producing more readable reports with provocative covers (as Sim suggests - Chapter 15), if they are to compete on equal terms with direct experience.

Evaluation cannot ensure that decisions are taken on educational grounds alone.

Logistics and politics play a big part. "No amount of evaluation ... will take the place of a dissemination base" (Reay) in securing adoption and some administrators will push ahead with a project, whatever the evaluation findings, as resources have been committed to it and their prestige is at stake. A wise evaluator tries to avoid providing information for a decision that has already been made, perhaps shifting to a related one which is still open.

Evaluators ought not to make decisions for the decision-maker.

Some decision-makers hope that they will but most evaluators believe that it ought not to be part of their job; the hope comes from a belief that statistics will lead automatically to clear-cut decisions. This is not the case, as decisions are seldom clear-cut. Decisions involve

balancing values. Evaluators have to make decisions at a different level in choosing techniques and date.

Evaluation will not necessarily make decision-making easier.

Evaluation challenges the fixed view by opening up ranges of things to be considered, of possible alternatives. This can make decision-making harder although hopefully the decisions will be better. Nevertheless, some decision-makers will welcome the clarification, the marshalling of pros and cons, which a well-structured report offers.

There is no agreed evaluation methodology.

The trend papers in this volume reveal a range of views on evaluation theory and the case studies a diversity of practices. It is difficult for the customer to be quite sure what he is buying!

A misconception of evaluation.

A common misconception of evaluation is to regard it as:

> 'essentially the process of determining to what extent the educational objectives are actually being realised by the program of curriculum and instruction. However, since educational objectives are essentially changes in human beings ... then evaluation is the process for determining the degree to which these changes in behaviour are actually taking place.'[1]

This definition is embodied in the Eight Year Study directed by Tyler in the 1930s.[2] Its dependence on behaviourist psychology and the psychometric tradition is apparent.

The influence of this definition has persisted over the intervening decades, although evaluations which over-relied on it have frequently failed in a number of ways. These are detailed in several places[3] and will merely be outlined here. Frequently little or no significant difference is found between the changes, cognitive and, particularly, affective, of experimental and control groups[4] (Reay); such differences which do occur do not provide useful formative information; it is impossible to control variables (cf, Reay), particularly class-room processes, to the extent the approach demands; it is difficult to accommodate informal data although these may influence the decision-maker strongly; the need to measure for achievement of objectives tends to emphasize those elements which are easiest to measure.

This style of evaluation, unable to provide data until the end of the experiment, is not flexible enough to respond to the changing information needs of a project team; it is not what Stake calls responsive evaluation.

As well as the failure of the achievement-of-objectives evaluation to provide satisfactory information, the curriculum model of which it is a part is increasingly regarded as inadequate.[5]

[1] R.W. Tyler, Basic Principles of Curriculum and Instruction, Chicago, University of Chicago Press, 1949, p. 105-6.

[2] F.R. Smith and R.W. Tyler, Appraising and Recording Student Progress, Vol. 3 of Adventure in American Education, New York, Harper, 1942.

[3] Stufflebeam et al, Educational Evaluation, op. cit.

Parlett and Hamilton in D.A. Tawney (ed), Curriculum Evaluation Today: Trends and Implications (Schools Council Research Studies), London, MacMillan, 1976.

[4] W. Harlen, Science: 5/13 - A Formative Evaluation, London, MacMillan, 1975.

[5] L. Stenhouse, An Introduction to Curriculum Research and Development, London, Heinemann, 1975, p. 70.

Innovations are seen by many as likely to be more successful if objectives are not over-detailed and pre-defined but allowed to develop during the project.

The scope of evaluation.

Most agree that evaluation needs to be wider than the measurement of achievement. Despite disagreement on which methodologies to adopt in its place, whether to develop more sophisticated statistical designs or to abandon the approach entirely, there is agreement on the scope of evaluation and on the areas needing investigation.

Context.

Many recent evaluations have included investigations into the context in which the curriculum project is working, into existing curricular and class-room practices, into alternative schemes which might be imported and adapted and into opinions on possible proposals. Such investigations are given a variety of names: reflective evaluation[1], presage evaluation (Sim), antecedent evaluation[2] (cf, Yoloye, Chapter 10).

Evaluators have looked at material resources such as laboratory apparatus; Sim describes such a study. Some have looked at the qualifications of the teachers involved. Other important contextual variables which may need investigation are the school timetable and the public examination system; Krasilchik refers to the difficulty of introducing a course which does not lead to the all-important university entrance examination. Sim also refers to the effects of a tradition of public examinations.

Evaluators seem to have been less ready to examine less tangible but perhaps fundamentally more important factors. Integrated science projects often advocate 'discovery methods' in which teachers act more as guides than authoritative sources of information; stepping down from the platform and across the boundaries of the disciplines may produce insecurity. At a more prosaic level, it may reduce prospects of promotion (Krasilchik, Chapter 14). Yoloye, looking at the stage of development of African primary schools, uses Beeby's classification to cast doubts on the compatibility of the African Primary Science Project with current teaching methods.

The problem is explored more thoroughly in an area with similar problems, namely the integration of the humanities, by Shipman[3] in a study which is of relevance to all curriculum developers, particularly those involved in cross-disciplinary projects.

How contextual information is used depends on the methodology adopted: in some the context is one source of the variables selected for measurement at the beginning of the study and also of some of the clues for interpretation of findings at the end; in others, the study of context extends throughout the study, its interaction with the innovation providing increasing insight.

Aims.

Traditionally, evaluators have been expected, as a pre-requisite to designing tests and schedules, to help their projects to clarify and operationalize objectives. More recently, with the increasing recognition that objectives emerge, develop and change (cf, Yoloye) throughout a project, they are likely to replace pre-definitions with a continual surveillance and attempt to clarify.

[1] D. Cohen, "Evaluation of Integrated Science Curricula", in Richmond (ed), New Trends, Vol. 2 op. cit., p. 143-65.

[2] R.E. Stake, "The countenance of educational evaluation", Teachers' College Record 68, Vol. 7, 1967, p. 538.

[3] M.D. Shipman, Inside a Curriculum Project, London, Methuen, 1974.

The aims of the curriculum project must be related to the wider context of the society of which the schools in which the innovation is to occur are a part. In countries with a well-defined society-orientated educational philosophy, a child-orientated curriculum may appear out of place (cf, Yoloye). In many, a wide range of views coexist; in consequence the aims of a project will represent a compromise and its organizers must expect its products to be used in a variety of ways.

Sim refers to parental anxieties which reflect a view that science is part of the serious business of acquiring qualifications; this suggests the need for those drawing up the aims of a project to be aware of vocational pressures; future employers and national manpower planners may well be influential.

At a more detailed level, Kelly[1] made interesting use of teacher ratings of the objectives of sixth form biology teaching, comparing their ratings of objectives for desirability and attainability. Discrepancies between these, and between these and actual attainment examined at different times in the life of the project, provided valuable formative data. Of particular interest as an example of responsive evaluation is how a discrepancy between the desirability and attainability of the objective 'handling quantitative information' led to an investigation which resulted in a considerable improvement in its attainment.

An aspect of integrated science which the evaluation may need to examine is the view of the nature of science reflected by a curriculum project, particularly through its aims. Two features make a careful examination of the view of science important: one is because an integrated science project frequently emphasizes process rather than content, the other because of its justification for integration, whatever is meant by this.

The problem arises in projects for older pupils when an attempt is made to teach a process by which science develops, the 'scientific method'. One model of this is "pattern-finding"; for example, in the Schools Council Integrated Science Project (SCISP), "pupils are asked to extract a pattern from the information"[2] gained from experiments they have performed. This process, called induction, has been discredited since the eighteenth century; for a review of the arguments against it, see Tawney, Jevons, or Medawar.[3] Although there is no consensus on how science progresses (except that it is not by induction!), it is a pity that more projects have not placed at least equal emphasis on a hypothetico-deductive approach in which pupils use their existing knowledge to suggest hypotheses which they then test experimentally. With this approach, it is easier to show, first, the essential part which existing knowledge plays in discovering new knowledge, secondly, that science is a human creation and, lastly, the strengths and limitations of science.

In a similar and closely related category is the rationale behind the notion of 'integration' used in a project. This has been explored elsewhere[4], but in suggesting this as an aspect of every project which may need evaluation, it is necessary to warn evaluators that some of the higher sounding reasons given for integration frequently do not appear to be embodied in the learning experiences advocated for reasons that are not always clear. Perhaps interviews with

[1] Kelly, op. cit., p. 95.

[2] W. Hall, Patterns: Teachers' Handbook, London, Longmans and Penquin, 1973, p. 30.

[3] D.A. Tawney, "The nature of science and scientific enquiry" in Sutton, C.R., Haysom, J.T. (eds), The Art of the Science Teacher, London, McGraw-Hill, 1974, p. 19-29; F.R. Jevons, Science Observed, London, George, Allen and Unwin, 1973; P.B. Medawar, "Hypotheses and Imagination", in The Art of the Soluble, London, Penguin, 1969, p. 147-73.

[4] J. Rutherford and M. Gardner, Integrated Science Teaching in New Trends, Vol. 1, 1971, op. cit., p. 47-55.

the project team may clarify this area and prevent resources being devoted to an empirical study of the achievement of aims whose statement was never meant to be more than educational rhetoric. Brown's analysis later in this volume of some affective objectives in the Scottish Integrated Science Project and the arguments given for including these is an example of one approach to this area.

Learning Processes.

What learning processes or transactions does a project advocate and how do they differ from traditional practices? And what practices do teachers ostensibly following a project, adopt? These questions are framed in terms of a curriculum project; there are corresponding ones for teachers evaluating their own work.

On what psychological theories are the experiences advocated by a project based? If it is adopting a 'discovery' approach, has it considered the objections to it (e.g. Ausubel[1])? Is the project team aware of the limitations of learning theories (e.g. those of Piaget and Gagne), indeed of social research in general[2]? Although the bulk of evaluation in this area will occur at the start of a project, evaluators must realize that attempting to implement learning processes will often give insights into the theories behind them.

To move from central policies out into the field, what are the teachers' understandings of a project's philosophy? For example, Sim describes a study which suggested that teacher conceptions of the term 'discovery' were limited.

What discrepancies exist between what teachers say they do and what they do? Increasingly, emphasis in evaluation is shifting to classroom observation, for there is agreement that the failure of curriculum projects to change science teaching by the extent hoped is due not so much to the failure of the classroom processes advocated as to their never really being practised. Jeffrey's report that "... many teachers were not attempting to teach towards these attitudinal objectives" is more generally true (cf, Yoloye).

This suggests that the failure lies not in curriculum strategy but in innovation strategy - the way a project interacts with teachers and indeed with the whole education system - to further the adoption of its curriculum strategy. The answer seems to be more and better teacher training.

It is an evaluation task to try by a study of process to separate the degrees of success of these two strategies; the failure of the measurement-of-achievements model to make the distinction was one reason why it was discredited. Evaluators will respond to the problem in different ways: some might obtain data using highly structured observation schedules and use sophisticated statistical techniques to relate the data to measured effects and other variables. Others might try to typify process by a case-study of a lesson and relate it qualitatively to effects. In no area of evaluation are the problems more difficult and the attempted solutions more diverse.

Effects.

The effects of an innovation cover a wider area than its objectives and are frequently unexpected; their measurement is even more difficult than that of achievement of objectives.

Integrated Science courses often precede other science courses and it is in these that some of the effects of the course may be detected. Science teachers in the United Kingdom who are considering adopting the 14-16 SCISP course are likely to be concerned with how pupils who have

[1] D.P. Ausubel, "Some psychological considerations in the objectives and design of an elementary-school Science program", Science Education, Vol. 47, No. 3, 1963.

[2] M.D. Shipman, The Limitations of Social Research, London, Longman, 1972.

followed this course will fare in the subsequent 16-18 courses leading to the important General Certificate of Education (G.C.E.) Advanced (A) Level examination, so that it is important to look for the effects of the new curriculum in this area.

There is equal need to monitor concurrent effects. An innovation which advocates informal relationships between teachers and students, which cast doubts on the value of formal note-taking and other traditions and encourages the questioning of authority, may bewilder pupils in a school with a traditional style of teaching and may cause sharp resentment from other teachers. However, the achievement of cognitive objectives is still an important aim of most projects and most evaluators will want to test for such achievement; the means they employ will depend very much on the assessment traditions of the country and, if there is a public examination system, the relationship of the course being evaluated to it.

Later in this volume, Reay refers to the value of a properly-designed test as a teaching tool, while Sim describes what happens when there is a poor match between the objectives of a course and a public examination.

For courses remote from a public examination and implemented in areas where testing is perhaps viewed with suspicion, other means of evaluation of achievement may be more appropriate: for example, the Diagnostic Statements developed by Harlen[1] for the "Science: 5/13 Project" represent a checklist of behaviours to enable teachers to monitor pupil progress. Evaluators differ in their views on the use of tests in this kind of context; some feel that reliability can be achieved only if properly standardized tests are used; others feel that tests specially designed to match project objectives will be more satisfactory, particularly for formative purposes, and will happily sacrifice reliability for unobtrusiveness and match between course and test.

Evaluation of affective objectives presents more of a problem. It is difficult to tell whether the failure of many existing tests to detect significant gains is due to the tests or to the failure of teaching to bring affective changes about. Certainly there seems to be a general tendency for pupil familiarity with a school subject to breed, if not contempt, a lessening of enthusiams for it[2] (cf, Jeffrey, Chapter 12). Attitude scales should be used with caution; some evaluators will prefer to replace them or complement their use with other measures such as interviews and classroom observations.

Evaluation Methodology.

To evaluate the areas indicated above demands a wide-ranging methodology and access to the viewpoints of a number of disciplines. Philosophy of science, psychology and sociology can all provide perspectives from which to try to order the complexity of an innovation in integrated science. It is clear that no single evaluator can possess an adequate knowledge of all these disciplines and therefore it will be necessary to call upon the services of sympathetic experts in order to present the project with relevant information. In doing this, it must be remembered that presenting too many data will hinder rather than aid decision-making; data must be structured into information.

The evaluator will need to make use of the whole range of evaluation techniques: tests, questionnaires, classroom observation, teacher diaries, interviews, meetings (cf, Krasilchik, Jeffrey). The degree of structure used in questionnaires and interview schedules will depend on conditions; for example, Yoloye reports that the African Primary Science Programme found Flander's observation schedules too complicated; thus simpler ones were developed. Resources have a strong influence and clearly a teacher evaluating a small curriculum development in a school is likely to use fewer measures than is a nationally-funded project. However, even well-funded projects have to count the cost; many evaluators would agree that pupil interviews

[1] Harlen, op. cit.

[2] ibid.

are a very informative and flexible technique, but it has been estimated that interviews may cost twenty times as much per pupil as a pencil-and-paper test. Teacher questionnaires are a cheap way of obtaining data but it is difficult to get a reasonable percentage returned and those which are returned have not always been critically completed, as Reay points out.

In addition to these techniques, the evaluator will want to use what Eraut[1] (in Tawney, 1976) calls intrinsic evaluation, that is, an examination of the project documents, statements of policy, records of meetings and materials to conduct an exploration of the assumptions, values and agreements which lie behind a project's proposals; Yoloye gives an example.

Evaluators disagree on the extent to which data provided by these means should be quantified. Particularly for a summative evaluation, some evaluators prefer a highly statistical approach (for example, multivariate analysis), in which all data are quantified and analysed to obtain the significant variables and to identify the relationship between them[2]. However, this is a cumbersome method and cannot provide information rapidly on the problems as they arise. For significant variables and relationships to emerge, the appropriate variables have had to be recognized from the start. Even if relationships are detected, their interpretations often remain highly speculative.

This author favours the more descriptive illuminative approach of Parlett and Hamilton which has its roots in social anthropology. This characteristically has three stages: 'investigators observe, inquire further and then seek to explain'. General observation of a project in action both at its base and in the field will suggest recurring trends and issues. Some of these will be selected as the subject of more sustained and intensive inquiry with observation more directed and systematic. The third stage consists in

> "... seeking general principles underlying the organization of the program; spotting patterns of cause and effect within its operation; and planning individual findings within a broader explanatory context".[3]

The flexibility of this approach enables the evaluation to focus on problems as they arise. Some evaluators may accept this but see on the debit side a considerable loss of rigour. As discussed earlier, whether or not traditional evaluation designs really have greater rigour is debatable, but certainly methods must be employed to make evaluation more rigorous than casual observation.

The methods used in the illuminative approach to obtain rigour are triangulation, re-iteration and exposure. By triangulation is meant the use of several methods to expose an issue or support an hypothesis. Student opinion revealed through interview may by itself be unreliable, but supported by the observations of teachers and the results of informal tests, may provide sufficiently reliable information. To employ other methods could be regarded as over-kill.

Re-iteration is the process of exposing early descriptions and interpretations of what is occurring to criticism by others and subsequently redrafting them for further exposure. Project newsletters or regular evaluation reports are one means of exposure and they also help to involve teachers in evaluation and to encourage an evaluation 'atmosphere' (cf, Reay).

Evaluation findings.

The different approaches of these two methods of evaluation are manifest in their reports. The report of a summative evaluation with a statistical emphasis is likely to say little about the informal feedback used for formative purposes; it will contain sections concerned with the <u>design</u> of the evaluation, the <u>analysis</u> of the data, the <u>findings</u> (weightings, correlations, degrees

[1] Tawney (ed), <u>Curriculum Evaluation Today ...</u>, op. cit.

[2] W.W. Cooley and P.R. Lohnes, <u>Multivariate Data Analysis</u>, New York, Wiley, 1971.

[3] Parlett and Hamilton, op. cit., p. 93.

of significance) and the interpretations of the findings. Whether or not statistically significant relations are established depends on the inclusion of the right variables at the start. Whether or not educationally significant information is obtained depends on other factors; too often the findings are either obvious or concerned with variables outside the control of the decision-maker. As was mentioned earlier, interpretations have often to be very tentative as they are made too late to be tested.

The report of an illuminative evaluation is less likely to distinguish between formative and summative evaluation. It will probably be chronological, as the collection of evidence and the interpretation of evidence occur simultaneously and these are closely linked with the resulting decisions.

The final report will rely heavily on reports issued during the life of the project in order to give the decision-maker a "... description of the natural history of our conclusions ... an opportunity to make his own judgements"[1]. Because there is seldom consensus on judgements, it is likely that the report will give several judgements from different participants in the project and from outsiders who have had contact with it.

This approach to evaluation is flexible but it is in practice difficult to obtain sufficient triangulation to achieve rigour; furthermore, although evaluators may publish their tentative hypotheses in a newsletter so that they are subject to critical scrutiny, they may get little response. The ability of this method to provide all the information needed depends on the sensitivity of the evaluators to issues as they arise and their ability to organize their resources so that they can respond to them.

Perhaps more important than the limitations of these two approaches is a lack of knowledge not so much of what data to collect and how to collect them but of how to present the decision-makers with balanced information in a form which will command their attention and influence their decisions as much as their selective but direct experiences of the Project. Some evaluators are experimenting with more and more arresting forms of presentation; the Malaysian Case Study presented by Sim gives an example.

Conclusions.

In practice, many evaluators occupy a position somewhere between the two extremes described, as exemplified by the case studies in this book. The case studies use a range of methods according to their resources and opportunities, aiming for Sim's 'continuous comprehensive, concerted and co-ordinated' 'Orchestra Model' but probably achieving his 'Jigsaw Model'. Hopefully there will not be too many missing pieces!

FURTHER READING

1. Curriculum Development generally and the place of Evaluation in it. Stenhouse (1975) reviews the field and advocates process and research models of curriculum development in place of the achievement of objectives model.

2. Science Teaching and 'The Scientific Method', Tawney, and the corresponding Science Teacher Education Project Topic, (eds. Sutton, C.R. and Haysom, J.T., London-McGraw Hill), The nature of science and scientific enquiry serve as an introduction.

3. Curriculum Evaluation. In Tawney (ed), Curriculum Evaluation Today: Trends and Implications (Schools Council Research Studies), London, MacMillan, 1974, the current position with particular but by no means exclusive reference to the United Kingdom, is reviewed and many of the ideas mentioned in this chapter taken further. For a review of science evaluation in the United Kingdom, Tawney, D.A. "Evaluation and science curriculum projects in the U.K.", Studies in Science Education, Vol. 3, 1976, p. 31-54.

[1] H. Becker, "Problems of inference and proof in participant observation", American Sociological Review, Vol. 23, 1958, p. 652-60.

3 Evaluation and decision-making in integrated science

Wayne W. Welch
University of Minnesota, United States

SUMMARY

Curriculum developers, teachers, and funding groups face a number of decisions concerning integrated science. Providing information to assist in these decisions is the function of evaluation. In this paper, several key decisions are identified and the evaluation literature pertaining to these decisions is reviewed.

INTRODUCTION

In recent years, there has been considerable interest in removing the traditional subject boundaries that have characterized science education. This movement has taken on a variety of names (integrated science, unified science, interdisciplinary, etc.) but the common thread has been a rejection of the traditional labels of chemistry, biology, and physics in developing new courses. In this chapter, the term "integrated science" will be used generally to represent this trend, although a more specific definition will be presented later.

The apparent growth of the movement is noteworthy. In 1968, only twenty-five integrated science projects were known to be in operation in junior and senior high schools in the United States of America. By 1973, the numbers had risen to seventy-three known projects in the United States and fifteen in other countries. By 1975, Showalter and his colleagues at Ohio State University were reporting 170 programmes throughout the world which could be called unified science programmes[1].

A different perspective on the degree of integration of science courses can be obtained from the "Ninth Report of the International Clearinghouse on Science and Mathematics Curricular Developments"[2]. This report summarizes 392 projects developed between 1956 and 1974. A total of 126 projects report their presentation approach as integrated science. No doubt the degree of integration varies considerably among projects. Nevertheless, the interest in the movement is clearly evident.

Thus, this represents an examinable development in international science education. Science educators involved in this movement find themselves facing a number of decisions concerning curriculum development, funding and programme usage. These decisions will require credible

[1] V.M. Showalter, "Rationale for an Unbounded Science Curriculum", School Science and Mathematics, Vol. 75, No. 1, 1975, p. 15-21.

[2] J.D. Lockard (ed), Science and Mathematics Curricular Developments Internationally 1956-1974: The Ninth Report of the International Clearinghouse on Science and Mathematics Curricular Developments. Maryland: The Commission on Science Education of the American Association for the Advancement of Science, and The Science Teaching Centre of the University of Maryland in College Park, 1975.

information obtained from careful and effective course evaluations. This requirement for information gives rise to the general problem of this chapter. What evaluative information is currently available to assist those facing a variety of decisions about integrated science education? The specific objectives of this chapter are as follows: (1) to define several key terms - evaluation, decision-making and integrated science; (2) to identify several important decisions regarding integrated science; (3) to review present evaluative evidence that pertains to these decisions; and (4) to offer several suggestions for future evaluation studies in integrated science.

DEFINITIONS

Evaluation

Evaluation is generally defined as the process of delineating, obtaining, and providing useful information for judging decision alternatives.[1] The use of the word judging suggests the important place of values in the evaluative process. The decision-maker is required to ascertain the relative value of competing alternatives. Evaluation is designed to assist in this valuing process by providing relevant, credible and valid information. Thus, evaluation is an information-generating process.

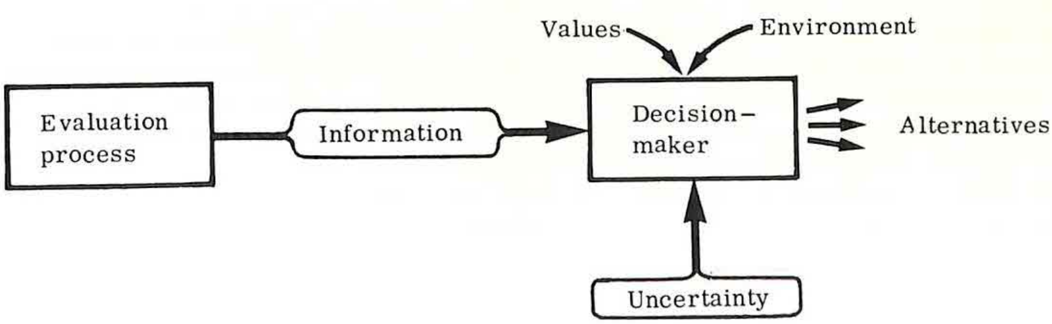
Decision-making

Decision-making can be considered in terms of both technology and human variables. The former stresses quantitative techniques such as those found in management science (e.g. programming, scheduling, simulation). These techniques can be used to reduce the degree of uncertainty in a decision situation. The latter concentrates on the problem-solving limitations of people. The decision-maker is viewed in terms of culture, personality, aspiration and problem solving strategy.

In this chapter, we will accept a simple view of decision-making where a decision-maker faces a known set of alternatives and selects a course of action by a rational selection process. However, we must remember that a decision-maker cannot recognize all feasible alternatives and, furthermore, is influenced by a biased perception of the environment. Thus, the actual cognitive process of deciding is indeed complex.

The key element in decision-making, as defined here, is choice among various courses of action. The major assumption we make is that information will improve the probability of a "successful" choice. Success is used in the sense of achieving some desired goal or objective. Whenever an individual is uncertain as to how best to pursue the objectives, indecision results. Reducing this indecision is the purpose of evaluation.

Fig. 2. Relationship between evaluation and decision-making.



The need for evaluative information arises from the uncertainty felt by the decision-maker. The choice of alternative courses of action is a function of the interaction of the decision-maker's environment, his value system, the nature of the uncertainty and available information.

[1] Stufflebeam <u>et al</u>, op. cit.

For example, suppose a teacher is considering changing the physics textbook. A number of alternatives are available: PSSC, Project Physics, Nuffield Physics, or a revised traditional text. The need for evaluative information regarding these courses is directly proportional to the uncertainty felt by the teacher. If the teacher is told what book to use, there is little uncertainty and no need for evaluation. If there is freedom of choice, greater evaluation information should increase the likelihood of a satisfactory choice. The values of the teacher (including such things as perception of the purpose of pre-college physics) as well as such personality characteristics as flexibility, need for change or conservatism and real and perceived environmental constraints (e.g. funds available, student ability, administrative pressure, or national examinations) also will influence the decision, and their existence must be recognized. The presentation of evaluative information about the various physics texts will reduce the indecision of the teacher and increase the probability of an effective choice.

In the curriculum area, there are several kinds of decisions to consider. Each decision type requires a specific evaluation focus to improve the probability of a successful decision. These types are shown in Figure 3.

Fig. 3. Curricular decisions and evaluation.

Type of decision	Major decision-maker	Evaluation required
Improvement	Course developer	Formative
Support of intervention	Funding group	Monitoring
Utilization or recognition	Course user	Summative

Formative evaluation addresses questions of course improvement during the development of a product. Monitoring evaluation is needed by the funding group (e.g. governments, publishers, or foundations) to answer the question of continuing support. Summative evaluation tries to provide information to potential course users (e.g. teachers) so they may make appropriate adoption decisions.

In the remainder of this chapter, we will review the available evaluation information that addresses each of the three decision areas in integrated science. But first it will be necessary to provide a more specific meaning of integrated science education.

Integrated science

Integrated science education has been advocated as a panacea for a multitude of science teaching ills. It has been offered as a way to increase scientific literacy, understand the processes of science, increase interest in science, meet learner needs, maintain flexibility and show the relationship of science and society. But these are the goals of nearly all science programmes and are not helpful in defining integrated science. An evaluative look at integrated science requires a description of its unique aspects in order to clarify how integrated science differs from non-integrated science. A review of the literature[1] yielded only five characteristics which

[1] Cohen, op. cit., T. Gadsden Jr., et al, "Cutting Boundaries with Correlated Science", School Science and Mathematics, Vol. 75, No. 1, p. 80-86; S.M. Haggis, "Attributes of Integrated Science Units", in Judith Reay (ed) Final Report of Integrated Science and Teacher Education (ISTE) Workshop, St. Augustine, Trinidad: University of West Indies, 1973; Showalter, "Rationale for ...", op. cit.; and I.L. Slesnick, "The Effectiveness of a Unified Science in the High School Curriculum", Journal of Research in Science Teaching, Vol. 1, p. 302-314, 1963.

seemed to distinguish integrated science from other programmes. Most other claimed characteristics would apply to any well designed science course. Figure 4 is a listing of the unique characteristics of integrated science.

Fig. 4. Characteristics of integrated science.

1. Traditional subject matter boundaries are phased out.

2. The course usually lasts for two or more years and is sequential.

3. The sequence tries to avoid duplication of content.

4. The course usually serves a general education function (i.e. to develop scientific literacy).

5. The course is organized around a selected unifying theme or topic.

These characteristics were the criteria used for selecting evaluative studies of integrated science in this chapter. These studies were examined to discover what information is currently available for the different decision-makers - developers, funding groups and teachers.

Formative evaluation results

Formative evaluation is usually carried out by course developers to assist in the decisions required for course improvement. Typical procedures include pilot testing materials in schools, reviews by teachers of preliminary versions of textbooks, and obtaining from scientists checks on the accuracy of the subject matter. Occasionally, achievement tests are used to determine if students are learning the intended content.

Most formative evaluation results are not published because of their low generalizability. However, the case studies in the present volume provide some good examples of the procedures (and problems) involved in producing information for curriculum improvement. (See particularly the chapters on the Caribbean Integrated Science Projects by Reay; the Scottish Integrated Science Syllabus by Jeffrey; the Integrated Science Teaching in Malaysia by Sim; and the African Primary Science Programme by Yoloye).

These chapters present several examples of how formative evaluation is usually conducted. Further examples of extensive formative evaluations have been reports by Welch[1] and by Cohen[2]. Cohen's chapter in Volume II of this series is particularly illuminating on two extensive formative evaluation plans - the Australian Science Education Project (ASEP) and the Schools Council Integrated Science Project (SCISP). The ASEP project is one of the few published examples in integrated science where the formative evaluation has influenced revision decisions. No doubt there are others but these are not well documented.

Whether or not the use of formative evaluation has improved an integrated science course is unknown. It seems apparent it has resulted in changes. It is likely these changes have been improvements but no direct test of this fact was uncovered in the review of literature. The research on the effectiveness of formative evaluation as a means of improving decision-making is virtually non-existent. This conclusion is not limited to integrated science but is characteristic of formative evaluations in various areas.

[1] W.W. Welch, "Curriculum Evaluation", Review of Educational Research, Vol. 39, p. 429-43, 1969.

[2] Cohen, op. cit.

Monitoring evaluation results.

No studies were found in the search of the published literature in the area of "monitoring evaluation". First, the term is not widely used even though it was proposed by Scriven[1] at the same time as his widely-used formative and summative terms. It refers to the generation of information for the supporting agency to reach decisions on the need for intervention and/or continued support.

The case study by Sim describes quite well the need for such information in addition to providing a few examples of how this information was used in Malaysia. This is one of the few published examples of how evaluation can respond to the need for monitoring decisions.

I am somewhat familiar with the monitoring efforts of the National Science Foundation, a large science funding agency in the United States. In general, very informal procedures are utilized - site visits, proposal review and required periodic reports. These are used for both integrated science and other courses. The monitoring procedures have come under recent attack because of potential bias or discrimination in funding decisions. Considerable effort is now being made to improve monitoring evaluation. The intent is to improve decisions on the intervention and support of curriculum projects. Of particular interest is the need for information to help decide if integrated science should be supported in a given country. Some assistance for this decision may be found in the summative evaluation studies reported next.

Summative evaluation results.

A substantial literature search was undertaken seeking summative studies of integrated science programmes. All issues of the United States published Science Education, School Science and Mathematics and the Journal of Research in Science Teaching were examined from 1970-1975. Any title found that might pertain to integrated science at the secondary level (ages 12-18) was reviewed and abstracted if it provided evaluation information. The review was limited to the secondary level because no effective way was found to separate integrated science from traditional science at the elementary level. A similar procedure was followed for the Australian-produced Science Education Research, issues 1971-1974, and Volumes 1 and 2 of Studies in Science Education from the University of Leeds. In addition, doctoral theses of the past twelve years were searched and six were found that had studied various integrated science programmes.

The purpose of this search was to identify those summative evaluation studies which provided specific information on integrated science programmes. A further restriction was that programmes meet the definition of integrated science given previously. Surprisingly, our search yielded only thirteen studies and only seven of these met the tenets of disciplined inquiry; namely - systematic, objective and founded on a test of evidence.

Three studies of programmes[2] perceived by some writers as integrated science were examined. However, the programmes were of one-year duration only and, therefore, failed to satisfy the definition of integrated science used in this chapter. The Klopfer-McCann and Cossman studies were of one-year unified science courses offered at university laboratory schools, while the Gardner study was of an interdisciplinary chemistry course. The first two studies found

[1] M. Scriven, "The Methodology of Evaluation" (1967), reprinted in Blaine R. Worthen and James R. Sanders (eds), Educational Evaluation: Theory and Practice, Worthington, Ohio, Charles A. Jones, Publishing Co., 1973.

[2] G.W. Cossman, "The Effects of a Course in Science and Culture for Secondary School Students", Journal of Research in Science Teaching, Vol. 6, No. 3, p. 274-83, 1969; and L.E. Klopfer and D.C. McCann, "Evaluation in Unified Science: Measuring the Effectiveness of the Natural Science Course at the University of Chicago High School", Science Education, Vol. 53, p. 155-64, 1969; and J. Gardner, "The Interdisciplinary Approaches to Chemistry (IAC) Program and Related Research" in R.P. Tisher (ed), Science Education: Research, 1973, University of Queensland, Brisbane, Queensland: Australian Science Education Research Association, 1973.

several positive effects on measures of scientific literacy (e.g. Test on Understanding Science), but both studies were specially developed courses rather than multi-year integrated science curricula.

Although I would be the first to admit that our search might not have included all papers, my initial reaction was one of surprise at the limited number of evaluation studies on such a popular movement. However, my goal was not number of studies but, rather, the important information contained in those studies. Unfortunately, my impressions did not change as the results of these studies were examined.

Results.

By far, the most popular target for evaluation studies has been the integration of chemistry and physics[1]. Furthermore, a number of similar studies was reviewed extensively in Johnson which dated back to 1938. However, these studies are outside the time frame of the current review (1963-1975). It should be noted, however, that nothing in the earlier studies is contrary to the conclusions reached by the author for the more recent period.

Although some may argue that the chemistry-physics integration is not unified science, it does meet the operational definition of integrated science presented in Figure 4. We are also fortunate to have several replications of a similar programme so that any conclusions reached are based on more than one integrated science sequence.

Each of the six studies had slight variations in content and in criterion measures. For example, Johnson used the Test on Understanding Science (TOUS) as his measure of science process, while Fiasca used the Watson-Glaser Test of Critical Thinking. Similar variances existed on the achievement measures, although the Co-operative Science Test was the most popular. A summary of the six studies is shown in Table 2, over page.

[1] C.B. Bundy, Comparison of Achievement in an Integrated Two-Year Chemistry-Physics Course with Achievement in Chemistry and Physics taught as Separate Courses. (Unpublished doctoral thesis; Indiana University, Bloomington, Indiana), 1969; M.A. Fiasca, Feasibility of the Integration of Selected Aspects of (CBA) Chemistry, (CHEMS) Chemistry, and (PSSC) Physics into a Two-Year Physical Science Sequence. (Unpublished doctoral thesis; Oregon State University, Eugene, Oregon), 1966; F.D. Goar et al, "The Moline Project", The Science Teacher, Vol. 33, p. 40-41, 1966; J.D. Jacobs, The Development and Evaluation of a Unified Chemistry-Physics Course Taught by a Cooperative Teaching Method for a Two-Year Period. (Unpublished doctoral thesis; Pennsylvania State University, University Park, Pennsylvania), 1967; G. Johnson, "An Integrated Two-Year Chemistry-Physics Course Compared with Consecutively Taught Separate Courses", Science Education, Vol. 59, No. 2, p. 143-54, 1972.

Table 2 reports the results of comparing the integrated physics-chemistry course with a two-year sequence of chemistry and physics taught separately. In all cases except Fiasca's study, the integrated course was locally developed and taught by the person conducting the study. All were two-year studies carried out at the upper secondary level (ages 15-18).

Table 2. Evaluation results for integrated physics-chemistry courses significant differences.

Study	Criterion tests				
	Chemistry achievement	Physics achievement	Physics-Chemistry achievement	Process of	Student interest
Jacobs	0	0	0	0	
Bundy	-	-	-		
Johnson*	+	0	-	0	+
Lerner	0	0			
Goar **	0	0			
Fiasca	0	0		0	0

Key:
+ integrated science higher
0 no difference
- integrated science lower

* Because I seriously questioned the appropriateness of the adjustment procedures, only unadjusted comparisons are reported here. The original study reports both.

** Statistical tests not reported. Percentile differences reported are not significant.

A total of twenty comparisons were made in the six studies. Two comparisons favored the integrated course, four showed the separate sciences higher, and fourteen resulted in no significant differences. The implications are clear. There is virtually no evaluation evidence in these studies to support integration of physics and chemistry over the teaching of the courses separately. It appears to make very little difference using the criteria of achievement, understanding the processes of science or student interest in science.

This conclusion is in sharp contrast with evaluation interpretations reported elsewhere[1]. In order to resolve this apparent difference of opinion, let us examine the other integrated science studies found in the literature search. Two studies[2] of the unified science programme at the now-defunct University School at Ohio State appear to have been particularly influential in providing evaluation evidence in support of integrated science. It is my contention that both studies have sufficient problems to warrant equivocal conclusions consistent with the integrated chemistry-physics findings. The Showalter study will be considered first because it is more recent and addresses a broader range of evaluation criteria. One-hundred-and-eight graduates of a unified science course from 1963-1965 were compared with one-hundred-and-eight 1960-1962

[1] Cohen, op. cit., Showalter, "Rationale for ...", op. cit.

[2] V.M. Showalter, Effects of a Unified Science Curriculum on High School Graduates. (Unpublished doctoral thesis: The Ohio State University, Columbus, Ohio), 1968, Slesnick, op. cit.

graduates of a traditional sequence at the same school. These are labelled "Experimental a" and "Control a" in Table 3. A further comparison was made between seventy-one matched pairs of the 1963-1965 graduates at Ohio State and similar graduates at a traditional Ohio school (called "Experimental b" and "Control b"). The groups were compared on a variety of measures including college grades, science interest, science literacy test, and questionnaire items. A summary of the findings is found in Table 3.

Table 3. Unified science versus traditional science** Test means.

Groups	Criterion measures						
	Science literacy test	Years of science	Hours of College science	Science interest	College preparation perception	College grades	Five questionnaire items
Experimental a	29.7	3.59*	7.18	12.3	19.01	2.22	No differences
Control a	27.4*/	2.77	6.06	11.3	17.42	1.92	
Experimental b	29.2	3.76*	7.02	12.7	19.26	2.29	No differences
Control b	31.4	3.20	7.31	11.8	18.08	2.46	

* Indicates a statistically significant difference.

/ Low score indicates higher science literacy.

** From Showalter (1968).

A total of twenty-two separate comparisons were made. Only two of them (years of secondary school science) favored the unified science group when statistical tests were applied. The control group was significantly higher only on the Science Literacy Test developed especially for this study. The remaining nineteen comparisons were not different. Thus, I believe the only justifiable conclusion is the one reached earlier. There is no evaluation evidence in this study to support the claims that integrated science achieves the objectives of science literacy better than the separate science approach. One apparent conclusion is that students who sign up for a four-year sequential course tend to stay in that sequence more than do those students who must decide upon a new science course each year (see column two).

The Slesnick study was an evaluation of the first three years of the Ohio State University School programme described above. The single criterion instrument was a measure of an individual's conception of the physical world using a test of a "rational image of the universe". A matched group was chosen from another school and seventy-eight matched pairs became the subjects of the study. The major finding was that the unified science students scored higher on this test than did the control group. Although there are methodological errors in the report (e.g. Table 4 on page 311 use of t-tests to test proportions and incorrect designation of sample size), my major concern is an assessment of the programme using a test developed by the same person who taught the experimental group. It also appears the test is content-specific; that is, it was designed particularly for the unique content of the unified course. This is hardly an unbiased measure of course outcomes.

These concerns do not imply that the results are incorrect but, rather, they must be viewed with extreme caution. Further verification is needed. Little was found in the remaining five studies to contradict this conclusion.

Hall[1] reported on an evaluation of SCISP developed for 13-16 year old pupils. He reports there were significant differences between the means of SCISP and non-SCISP pupils of the same age. However, he does not report which group scored higher. Furthermore, other aspects of the study are not clear, e.g. control group characteristics, statistical tests used and group means.

Several of the studies reported the positive reactions of the integrated science experiences as perceived by the teachers involved in teaching the programme. In many instances, evaluators reported that the programme seemed to work well but did not provide supporting data. The one exception to this general rule was in the area of student enrolments. Herr[2], Pfeiffer[3], Fiasca and Gadsden all report increased science enrolments after the implementation of an integrated science curriculum.

Whether or not this represents a true expression of student reaction to the movement, or is merely an artefact of sequencing or the "halo" effect, is unknown. Information on this issue is currently not available. Certainly, it is an area which requires further study.

Conclusions and Recommendations

The purpose of this chapter has been to examine the information available to assist developers, funding groups and teachers in making decisions about integrated science. Evaluation and decision-making were defined and several essential characteristics of the integrated science movement were identified. A search of the literature of the past dozen years revealed few studies bearing on the problem. In the opinion of the author, what is available provides very little evidence to answer either utilization or support decisions. Furthermore, the value is unknown of the formative evaluation which has been conducted. Descriptions of projects are available and the number of integrated science programme throughout the world suggests a great interest in the movement. Unfortunately, the situation is little changed from 1969 when this reviewer examined 204 documents in science education. At that time I wrote, "a finding surprising to this reviewer was the small number of studies devoted to unified science. Recent activity suggests a movement in this direction. The fact that only four studies reported during a two-year period (1968-1969) were concerned with this topic indicated little research interest in the movement[4]". (Welch, 1972, p. 98). A similar situation appears to exist today.

Given the lack of valid and credible information and the need for some key decisions regarding integrated science, I would strongly urge the planning and initiation of several well designed evaluations of the more popular programmes in existence today. Longitudinal studies, perhaps funded by international agencies or national agencies such as the Nuffield Foundation or the National Science Foundation and carried out by impartial evaluators, are sorely needed. Philosophically and psychologically, a strong case can be made for the place of integrated science in our curriculum. A valid test of this rationale is needed by teachers, the public, and the funding agencies. Information is required to make effective and efficient decisions regarding the role of integrated science in our schools.

[1] W.C. Hall, "A Description of the Development of Assessment Procedures for the Schools Council Integrated Science Project", in Fisher, op. cit.

[2] L.G. Herr, "Unified Science: A solution to Physics Enrolment", The Physics Teacher, Vol. 9, p. 248-52, 1971.

[3] C.H. Pfeiffer, The Development and Implementation of a Four-Year Unified, Concept-Centered Science Curriculum for Secondary Schools, Final Report. ERIC Document ED 054 965 SE 021 468. (ERIC Document Reproduction Service, P.O. Box 190 Arlington, Virginia 22210, U.S.A.), 1969.

[4] W.W. Welch, "Review of Research 1968-69 in Secondary Level Science", Journal of Research in Science Teaching, Vol. 9, No. 2, p. 97-122, 1972, p. 98.

ANNOTATED BIBLIOGRAPHY

Cohen, D., 1973, "Evaluation of Integrated Science Curricula" in New Trends in Integrated Science Teaching: Volume II (Ed. Peter Richmond), Paris, Unesco.

A chapter in an early volume of this series that examined the evaluation needs of integrated science. Interest in this chapter led to the current volume.

Haggis, S.M., 1973, "Attributes of Integrated Science Units", in Final Report of Integrated Science and Teacher Education (ISTE) Workshop (Ed. Judith Reay), St. Augustine, Trinidad: University of West Indies.

Characteristics of integrated science are examined in this report of a workshop devoted to the topic.

Lockard, J.D. (Ed), 1975, Science and Mathematics Curricular Developments Internationally 1956-1974: The Ninth Report of the International Clearinghouse on Science and Mathematics Curricular Developments. Maryland: The Commission on Science Education of the American Association for the Advancement of Science, and The Science Teaching Centre of the University of Maryland in College Park.

An excellent summary of curriculum projects around the world. Volumes in this series appear every two years.

Scriven, M., 1973, "The Methodology of Evaluation", reprinted in Educational Evaluation: Theory and Practice (Blaine R. Worthen and James R. Sanders, eds), Worthington, Ohio: Charles A. Jones Publishing Co.

Scriven is a leading evaluation theorist. His methodology paper has had considerable influence in the profession and has been reprinted many times. The Worthen and Sanders book is a very good general text, available in paperback.

Showalter, V.M., 1975, "Rationale for an Unbounded Science Curriculum", School Science and Mathematics, Vol. 75, No. 1, p. 15-21.

The complete volume of this issue is devoted to unified science. Particularly useful in defining integrated science. Additional data on integrated science usage can be obtained from FUSE, The Ohio State University, Columbus, Ohio 43210, U.S.A.

Slesnick, I.L., 1963, "The Effectiveness of a Unified Science in the High School Curriculum", Journal of Research in Science Teaching, 1, p. 302-314.

One of the earliest studies of integrated science. Includes a rationale for the movement and defines key characteristics.

Stufflebeam, D.L. et al, 1971, Educational Evaluation and Decision Making, Itasca, Illinois, F.E. Peacock Publishers, Inc.

Perhaps the best resource book on the relationship between evaluation and decision making.

Welch, W.W., 1969, "Curriculum Evaluation", Review of Educational Research, 39, p. 429-43.

A review of curriculum evaluation results of the 1960s. Definitions of terms are given and some assessment of the quality of the evaluation is provided.

Welch, W.W., 1972, "Review of Research 1968-69 in Secondary Level Science", Journal of Research in Science Teaching, Vol. 9, No. 2, p. 97-122.

One of the reviews published periodically by the Educational Resources Information Centre (ERIC) at The Ohio State University, Columbus, Ohio 43210, U.S.A.

FURTHER READINGS

Schools Council Research Studies, (1973) Evaluation in Curriculum Development: Twelve Case Studies. London, Macmillan Education Ltd.

A series of papers by the evaluators of several Schools Council projects. Several are of integrated science projects. The chapter by Barry MacDonald (Chapter 7) addresses most directly the relationship between evaluation and decision making.

4 The analysis of integrated science curriculum materials

Jan R. Bloch
IPN on der University, Kiel, Federal Republic of Germany

SUMMARY

Three functions of the analysis of curriculum materials are explored: training, research and evaluation. Emphasis is placed on the latter in relation to the goals, the setting and the role of the evaluator. Formalized schemes for analysis, created in the United States, the United Kingdom and the Federal Republic of Germany, are presented and reviewed in terms of the criteria for their construction and their use in analysing integrated science curricula. The concluding section of the paper considers the relationship of analysis and analysis schemes to integrated curricula in respect to both content and methodological considerations.

The analysis of curriculum materials

The functions of an analysis

The analysis of completed curricula materials, either student materials or teacher materials or both, using one or another of the formal schemes described below, can have several functions.

The analysis itself can have a learning function. A teacher or a student teacher can learn much about the components and structure of a curriculum by analysing it. The analysis of a curriculum, or parts of several, should make pre-service or in-service discussions about curricula more realistic, as well as providing detailed information about curricula analysed. By using a formal scheme as a guideline for analysis, much information can be gained. Such pre-structured schemes limit subjective critiques or emphases on one area or another by separate analysts. Also, however, such a scheme necessitates that the person using it is trained in its use, a beneficial pre- or in-service situation.

An analysis of curriculum materials can also service a research function. The authors of the Sussex scheme designated "this as a curriculum criticism" function:

> " ... we prefer to use the term 'curriculum criticism' to describe a curriculum analysis which is not specifically decision-oriented; and see the main purpose of such criticism as the disclosure of meaning and the extension of knowledge about the curriculum. The critic, unlike the evaluator, is free to choose his own standards and values and to focus on particular issues rather than attempt to cover a wide range. Moreover, it is he, together with other educational researchers, who should provide the basic knowledge on which the evaluator can draw. The evaluator is essentially a technologist and his service role depends on the state of educational knowledge".[1]

[1] M. Eraut, et al, The Analysis of Curriculum Materials, Brighton, Sussex, University of Sussex, Occasion Paper 2, 1975.

Here the emphasis could be on extending knowledge for curriculum developers, for example, on what is available and what it is like with regard to structure, content and integrating principles.

An analysis of curriculum materials can also have an evaluation function. Here we must very carefully set out what is meant by evaluation and for whom. There comes a time when a teacher or a supervisor or an administrator wants to recommend for usage a curriculum which suits his own needs, attitudes, background or whatever. Such recommendations should be justified against "acceptable" or "desired" criteria.

The evaluator engaged in an external analysis of curriculum materials is placed in the position of decision-making. This he can do principally on two levels: by orienting his interest to internal or external relevance. Internal relevance indicates how the set goals are met by the concrete curriculum documents or how intracurricular, axiomatic agreed-upon requirements fit with the programme in question. It comprises an estimation of how the documentary and empirical evidence is consistent with respect to supra-ordinate goals and theoretical statements within the document or how representative the data are for a given concept. Internal relevance is akin to evaluating an industrial product with respect to its initial specifications or to generally accepted standards. Quality control and assessment of the product and its performance remain within the scope of agreed-upon levels, factual needs and accepted specifications.

This pre-supposes a given value-system, within which refinements are related to the accepted standard goal. The assessment of how well the final product meets the initial aim operates on the operationalized level of both specifications and data: reproducible, quantifiable, describable, unambiguous.

The internal relevance approach is valuable for documentation purposes and serves as a basis for information. It suffers at the same time the danger of triviality, as Smith warned:

> "It seems ... that a remarkably high proportion of the research reported is clean, stringently conceived, and effectively executed, reflective of rigorous and painstaking thought and experimentation, and remarkably trivial! There are occasions when I have the unpleasant fantasy that psychology has become so enamored of method that techniques become our independent variables and our substantive problems only the dependent ones". [1]

External relevance indicates how the given document and its implicit goals relate to external requirements. This implies: (1) sensitivity towards trends and changes within the value system, (2) anticipation and sketching in outline the features of foreseeable needs and goals; and (3) the evaluation of the theoretical frame of a curriculum and the evaluation of its importance. External relevance seeks to determine the discrepancies between the curriculum goals and the external requirements. The criteria beyond the scope of the curriculum system can be derived from theoretical deductions of societal requirements, the research interest of the evaluator, the reflected intentions towards a different society and the factual evidence of disagreement and discrepance between the theoretical system and the actual requirements.

By analogy to industrial products, in the case of external relevance, the product will not be subjected to quality and performance control with respect to the initial specifications, but the specifications will, first of all, be the subject of an evaluation of their relation to changed and changing value systems. After this kind of assessment, the feasibility of internal relevance of a curriculum gains importance.

Decisions based upon internal relevance use criteria often restricted to a descriptive analysis of the documentary evidence of the curriculum materials or empirical evidence from responses to the curriculum. The evaluator might claim neutrality by working within the system, but this is debatable.

[1] M.B. Smith, Editorial, Journal of Abnormal Social Psychology, No. 63, 1961, p. 461-65.

In the introduction to the Sussex-scheme The Analysis of Curriculum Materials, the authors enter the debate about the evaluator's role as being more than simply about that of description.[1]

They contrast Scriven's assertion that the evaluator must judge, that curriculum evaluation should involve comparison between alternatives and lead to definite conclusions, with other viewpoints which maintain that the evaluator should stop short of passing final judgements. The basis of this view is that different people in different contexts have different standards and different values. The Sussex authors contend that the evaluation should respect these values and provide evidence (documentary or empirical) for helping decision-making groups to make their own decisions on their own terms.

Goal- and value-guided research enters the field of doubt: for instance, are the priorities set for certain objectives sought by the evaluator based on personal preference or on a valid reasoning which reflects "objective" and "true" need? Are the criteria for judgement of extrinsic elements (that is, those elements that do not relate to intracurricular features such as consistency) gained by a constant process of "responsive evaluation" (Stake), by general consensus of a general community, or do the criteria stem from a theoretical analysis of the society and a prognostic synthesis of a desired society? It can be stated that these value concepts are not problems of "personal preference", that a common denominator which reflects intersubjective validity exists. It is well known that the assessment of judgement data is difficult. Scriven claims: "... value judgements are one of the slipperiest species in the whole logical zoo"[2], while Stake describes judgement data as messy[3].

The task of assessment is related to the heterogeneity of values. Stake remarked that "the evaluator's task is to reduce the apparent heterogeneity to a manageable representation ... in a simple yet valid way".[4] This recognizes the priority of value judgements, the secondary position of the methodology, and the provision and analytical treatment of judgement data (see also the discussion of decision-making by Welch in the previous chapter). The role of the judgement data within the process of decision-making might be viewed as providing guidelines for the evaluator to reach decisions between various curriculum programmes or in support of the construction of a new programme to meet his standards. The study takes its direction from the investigators commitments, belief, political position or, in many important cases, hunches.

Intentionality of the evaluator can be seen as an important component. The context of intentionality operates within the anticipation of not-yet-established goals, within the concept of "from Is to Ought". Intentionality also includes stated external criteria which relate to a value system but, at the same time takes advantage of data provided by descriptive analysis along with the methodology of assessing these data. The role of intentionality is one of the central themes within curriculum analysis systems and plays a key role in the description-evaluation debate.

This problem of "two cultures" is also recognized by the authors of the Sussex scheme:

> "Analysing materials is an obvious and convenient form of educational enquiry; and we expected to find considerable support in the literature. However, whenever we looked for guidance, whether it was to philosophy, psychology, to curriculum theory, to sociology, or even to practical books on the methodology of particular subjects, we seemed to find a large gap between what was written and what was needed for our analysis. We found plenty of theorising about what ought to be done and plenty of practical advice on what to do:

[1] Eraut et al., op. cit., p. 12.

[2] M. Scriven, Value Claims in the Social Sciences, Boulder, Colorado: Social Science Education Consortium, Publication No. 123, 1966, p. 1.

[3] R.E. Stake, Objectives, Priorities, and other Judgement Data. Review of Educational Research (Boulder, Colorado), Vol. 40, No. 2, 1970, p. 182.

[4] ibid., p. 184.

but there were no links between the two and there was little or no critical examination of the assumptions underlying practice. We had to conclude that, with a few notable exceptions, publications that might be appropriately labelled 'curriculum criticism' do not exist". [1]

Their work in curriculum analysis has led the Sussex analysts to believe that "many decisions in the field of curriculum are based on intuitively held unformalized knowledge", therefore, "the importance of curriculum criticism is hard to deny. It is needed both to illuminate existing knowledge and as a springboard for the heuristic leaps of the future". [2]

Curriculum Analysis Schemes

In this section, three of the available analysis schemes will be examined, together with the criteria to which they do, or do not, attend.

The Curriculum Materials Analysis System (CMAS), developed by the Social Science Education Consortium (SSEC), in Boulder, Colorado, was one of the first and in its longer forms, is still one of the most comprehensive of the analysis schemes. It exists in three versions - short, intermediate and long - each intended to have different purposes. The long form, offering a complete, in-depth analysis, is intended for use by students and professors in methods courses at university. The intermediate form is less detailed and offers a format more useful to curriculum committees and for in-service training. The short version is designed to give a brief overview of the curriculum to aid curriculum committees, supervisors, etc., in acquiring information prior to curriculum selection.

The impetus for the creation of the CMAS was the increase in number of available curricula in the social sciences in the United States in the 1960s. The creators of the scheme both developed the scheme and carried out analyses of various social studies curricula which resulted in the production of several reports, called CMA's which stands for Curriculum Material Analysis. (Both the scheme in its three forms and the various CMA's and other supporting documents are available from the SSEC headquarters in Boulder, Colorado.)

The intermediate version of the scheme (SSEC, 1971) will be used here as a basis for further clarification. This version has eight sections. They are:

1.0 Product characteristics.
2.0 Rationale and objectives.
3.0 Content.
4.0 Theory and strategies.
5.0 Antecedent conditions.
6.0 Evaluation.
7.0 Background of materials development.
8.0 Background of the analysis.

Each section is divided into two to six headings; each heading may be further divided into as many as six sub-headings. For example, section "3.0 Content" is divided into:

3.1 Cognitive content.
3.11 Author's view of subject.
3.12 Cognitive content of curriculum materials.
3.2 Affective content.
3.21 Author's view of affective content.
3.22 Affective content in the curriculum materials.

[1] Eraut et al, op. cit., p. 14.

[2] ibid., p. 15.

The section begins with a short clarification of what is meant by both cognitive and affective content. The material also contains definitions of such terms as "fact", "concept", "value", and "attitude" in order to ease the task of the analysts. For each of the headings and sub-headings, there is a number of questions: for example, there are eight questions for heading 3.1, two questions for sub-heading 3.11 and six for sub-heading 3.12. In total, there are twenty-four questions to be answered in section 3. Both the curriculum materials and supporting documents must be surveyed in order to answer all the questions.

The questions are of three general types:

1. Most of the questions are of a semantic differential type, for example 3.1, question 7: "Do the materials emphasize the affective or cognitive content?"

/ 0 / 1 / 2 / 3 / 4 / 5 / 6 /

A great deal of affective content	A balance	A great deal of cognitive content

2. A few questions are of a check list type, for example 3.1, question 2: "What discipline(s) is(are) emphasized in the materials?"

Anthropology	
Economics	
Geography	
History	
Political Science	
Psychology	
Sociology	
Social Psychology	
Interdisciplinary	
Multidisciplinary	

3. There are occasional open answer type questions, for example 3.1, question 3: "What other subject areas are emphasized?"

The total intermediate form has 376 separate questions to be answered. No estimate is given of the approximate completion time.

The analysis of the system is by no means complete, but nevertheless we can make observations about the scheme. First, it must be pointed out that the scheme contains certain definitions, for instance, of "values" (a value is assessed worth toward a thing, event, behaviour or phenomenon, 3.2, question 2), which are to be used by the analyst during the analysis. Assuming that these definitions are neither too "conservative" nor too "radical", an analyst should be able to maintain the scheme's stance during the work. Of the schemes presented here, this scheme is unique in that it establishes certain criteria, that is, meanings for terms.

The constructors have selected definitions which will cause as little disagreement as possible, whereas most other scheme constructors have left the question of criteria to the audience/analyst and thus maintained a neutral stance within the scheme itself.

The SSEC scheme was not developed with the natural sciences in mind. However, it is readily adaptable to almost any subject field, with the advantage that its emphasis in the areas of values and the affective domain is quite relevant. Should such a content section be developed for the natural sciences, it would be preferable not to use the semantic differential type question.

By far the largest difficulty with the SSEC scheme is its complexity. The constructors seem to have realized this themselves, in that they set out to do analyses of various curricula using the long form. Even the intermediate form, for one not carefully trained in the subject matter and in the use of the scheme, is difficult and time-consuming. Unfortunately, the short form is considered by many to be too short to be useful.

The SSEC, in its work in constructing the three versions with different audiences in mind and in its function as a clearing house for information about other such evaluation schemes in the social sciences area (they list some 800 schemes with different focus points) has promoted the notion of evaluation of curriculum materials. The SSEC scheme itself is a good example to those concerned in evaluation of science curriculum materials. It promotes the notion that adaptation of existing schemes rather than creation of others might be the way to approach the problem.

A second analysis scheme, The Sussex scheme for the analysis of curriculum materials, developed at the University of Sussex, is not subject area dependent. Three advantages are claimed for usage of this formalized scheme for analysis of curriculum materials, namely: (1) provided the analyst understands the scheme, most of the major areas are covered; (2) it is useful when two or more competing materials have been analysed with the same scheme; and (3) the scheme provides a structure for the analyst's arguments and clarifies the assumptions and evidence base.

Some general disadvantages of using such a scheme include: (1) the scheme must be one that any analyst can "make his own" without feeling either too free or too structured; (2) the danger of "overkill", that is, of spending too much time and effort in the analysis, of being sidetracked from the over-all goal by too much attention to detail; and (3) the communication problem, i.e. in what way and with what detail the report of the analysis should be disseminated.

The authors of the Sussex scheme describe three stages within the total evaluation of the curriculum at which the analysis component has different roles and different audiences, but with the same goal, namely, "to analyse all the available evidence and relate it to different educational perspectives".[1] (See Table 4.)

Table 4. The three stages of curriculum evaluation.

Time	Audience	Purpose
1. Formative evaluation	Development team	To guide further work
2. Initial stage of summative evaluation	Evaluator herself/himself	To guide the subsequent stages of the evaluation
3. Final stage of summative evaluation	Decision-makers	To guide their decisions

The authors suggest that during the formative stage, the analysis of the early written materials of a curriculum development project should be analysed in terms of major assumptions about feasibility, desirability and educational value, as well as goal analysis, consistency analysis and contingency analysis. They suggest that content analysis might be beneficial and that a cost/benefit analysis, although difficult at this stage, might yield sufficient important information.

[1] Eraut et al, op. cit., p. 13.

At the second stage - the initial stage of summative evaluation - the analysis is seen as having the function of hypothesis-formation for the evaluator as she/he guides and/or plans subsequent empirical investigations. Without such an analysis at this stage (pointing out possible incongruencies between the intended and observed antecedents, transactions and outcomes), the authors suggest that data-collection by the evaluators could be misdirected.

In the third stage - the final stage of summative evaluation - the analysis has as its role the combining of both documentary and empirical evidence into a concise final report which aims to inform decision-makers about the adoption or implementation of the curriculum.

The authors of the Sussex scheme see the functions of such a scheme falling into several categories, not just evaluation. Three of these functions are described below:

1. Curriculum criticism function: here the analysis is not specifically decision-oriented, but rather has, as its main purpose, the "disclosure of meaning and the extension of knowledge about the curriculum".[1]

2. Pre-service education function: depending upon where such an analysis scheme is used in pre-service education, its function could be quite different. The authors suggest that the use of such a scheme in a professional course to compare and contrast two rival curricula, or in a philosophy course to emphasize contingency analysis, can be beneficial. In this sense, a curriculum analysis approach can have an integrative function in pre-service education, but without careful planning a fragmentation can occur which would not be positive.

3. In-service education: the Sussex scheme arose when the authors were active in training teachers and others for curriculum work at the local level. Their experience suggested that an emphasis upon curriculum analysis has been successful in imparting the notions about, for example, what a curriculum is. Likewise, the authors outline three goals for short-term (one week) in-service curriculum analysis workshops for which they claim success. These are: to improve the implementation of new curriculum materials, to improve existing curricula and to guide the selection of curriculum materials.

The Sussex scheme itself is divided into four required and one optional part. The four required parts are: (1) introduction, (2) description and analysis of materials, (3) the materials in use, and (4) evaluation. Part 5 (Decision-making in a specific context) is optional because it focusses on constraints, patterns of use and implementation of strategies for a particular situation.

Part 1 (Introduction) has three major divisions:

1.1 Basic facts.
1.2 Author's rationale.
1.3 Issues and perspectives.

Section 1.1 involves a description of the title, author, publisher and cost of each of the parts of the curriculum, a statement of the aims and functions of the material, a description of the audience (age, school type, course duration, etc.), information about field testing (if any) of the curriculum, and additional information about the authors' credentials, background, etc. Both the curriculum material and supplemental materials must be used to complete this part of the analysis.

[1] Eraut et al, op. cit., p. 23.

Each of the five parts is subdivided in a similar way, with specific areas mentioned for comment. For example, in Part 3 (The materials in use) one finds the questions:

3.4.1 How much teacher time is needed prior to implementation for activities such as gaining familiarity with the curriculum, further planning, and selecting or producing further materials?

A question from Part 4 (Evaluation) is as follows:

4.1.3 What information about the users of the resource and their experience is available?

The scheme is both structured and flexible, in that the questions are prescribed and yet do not limit inclusion of further areas of analysis which might be relevant for a particular curriculum. Likewise, the scheme is flexible in that there is no set mode of answering. (For a further discussion of flexible evaluation schemes, see Tawney's discussion in Chapter 1.) This, of course, limits the kind of reliability investigations that one can make about the scheme itself. The authors stress that training in the use of the scheme in one-week workshops is useful, if not necessary. This could decrease the usefulness of the scheme because of the lack of training personnel for such a workshop and because of the lack of attractiveness of such an activity for teachers and supervisors. It is estimated that a thorough analysis using all secondary materials, especially for Part 4 (Evaluation), would require one to two weeks of full-time work. Outside of activities in pre- or in-service training, this time seems long for the experience and/or information gained.

Because the Sussex scheme was intentionally developed to be applicable to the analysis of any subject area curriculum for any age level, it lacks certain characteristics (for instance, any mention of integrating principles) which would make it more relevant to the evaluation of integrated science curricular materials. However, the bases on which the scheme was developed and author considerations of the functions and advantages/disadvantages of such a scheme make its consideration worthwhile.

A third formalized scheme is the Curriculum Material Analysis Scheme for Science (CMAS) developed at the IPN in Kiel, the Federal Republic of Germany. This scheme, as its title implies, was constructed specifically so that curricula in the natural sciences could be analysed. The authors contend that there are two groups which could benefit from systematic analysis of curriculum materials: curriculum constructors and researchers in curriculum design (to these groups might well be added decision-makers responsible for adoption of curricula).

The five points of emphasis in the construction of this scheme were:

1. Adaptability: it should be possible to later add questions not formally considered during the design of the instrument without changing the pre-existing structure of the instrument.

2. Specificity: information should be provided on more than one level.

3. Reliability: the use of the instrument by separate analysts on the same material should lead to similar results.

4. Quick access to the data: the data compiled should be quickly, succinctly and continually available.

5. Ease of use: little or no training beyond an "Instruction for the user" manual should be needed in order to use the system reliably.

Various structural features of the system have been incorporated to meet these five requirements. For instance, each of the various categories of analysis developed stands by itself

and gives its own information. However, by differentially combining the categories, other questions can be asked and answered. The use of a punch-card system for data storage and retrieval helps fulfil the requirements for adaptability, quick access and ease of use.

Table 5. Five major areas with one or more prime components.

Major area	Prime components
Content	Behaviourial elements; subject matter elements; general elements.
Instructional methods	Management; learning styles; educational technology and media.
Adaptiveness	Individualization; learning rate; integration.
Effectiveness	Evaluation techniques used in the curriculum; evaluation studies on the curriculum.
Administration	Cost; implementation; teacher training.

Because the CMAS was especially developed for analysis of science curricula materials, a prime area of interest is the subject matter analysis. The designers of the CMAS had eight requirements in mind when they constructed this part of the scheme, related to such factors as differential level of specificity, approach, comprehensibility, flexibility and cross-science references. These considerations led to the construction of four categories to indicate content. They are:

Column 45: Major area of subject matter

0 = biology
1 = chemistry
2 = physics
3 = geology or meteorology
4 = astronomy
5 = mathematics
6 = general to science

Column 46: General approach to subject matter

0 = none of the following
1 = biological
2 = chemical
3 = physical
4 = sociological
5 = economical
6 = technological
7 = historical or biographical
8 = mathematical
9 = environmental

Columns 47-49: Specific keywords

Each keyword is indicated in two lists, a general alphabetical list and a subject-oriented list (for biology, chemistry, physics, geology, astronomy, mathematics and general to science).

Column 50: General keyword

0 = none of the following
1 = measurement of
2 = forms of
3 = structure of
4 = application or use of
5 = instruments in the field of

Using these six columns, one can construct sentences describing a section of the curriculum as to content. For instance, if Column 45 = biology, Column 46 = chemical, Columns 47-49 = DNA and Column 50 = structure of, the description of the content could read: "a biochemical approach to the structure of DNA".

The use of columns 45 and 46 help in a limited fashion to describe whatever integration there might be of content in the curriculum. Likewise, emphasis in Section II of the CMAS concerning the total curriculum (not individual sections) is placed on analysing the principles used to "co-ordinate or integrate various science disciplines".[1]

The CMAS is one of the few schemes whose reliability has been tested. The constructors report a study in which fifty of the categories were analysed. In forty-seven of these fifty categories the inter-analyst consistency was acceptable at the 5 per cent level. The three categories where the consistency was not acceptable were in the knowledge categories, where distinctions are not clear to a non-specialist.

The CMAS is structured so that one does indeed learn a great deal about particular randomly-chosen parts of a curriculum (in the Section I analysis) and an overview of the total curriculum (in the Section II analysis). Perhaps one feels more secure about using this system in the analysis of science materials because it was developed especially to do that. However, outside of the subject matter analysis, it may well be that one of the more general schemes could offer more comprehensive information about the total curriculum process. What none of these schemes, nor any other known scheme does, is allow us to evaluate, or even analyse, the methods or extent of integration in science curriculum materials.

The relationship of analysis-evaluation schemes to selected bases for the integration of science.

The nature and the analysis of integrated science curriculum materials can be linked in a productive inter-relationship in which the analysis scheme stretches the curriculum criticism function towards actual curriculum construction.

If we picture the situation in which the analyst faces the question: "Does a curriculum exist that meets my criteria?", a first systematic search might consist of seeking an analysis scheme that matches the criteria. This is the first confrontation of initial value judgements with the schemes. If a scheme is not located, the list of criteria might be used to develop an evaluation scheme based upon a set of stated requirements to be met by an instructional programme

These requirements would not derive from common practice or traditional intentions within curriculum construction, but from deductions, for instance from the field of social theory, philosophy or philosophy of science and would be related to the objective within the concept of what has not yet been and needs to be established. This analysis scheme can contain elements which are not sufficiently or at all represented in existing science curricula with an integrated approach. The scheme gains, then, the normative status of postulating the need for these elements and can thus serve as the backbone for a curriculum to be constructed.

The second basic question of the analyst could be: "What are curricula like that call themselves integrated?". Here the focus is on the various principles of integration: the scheme should be able to speak to these principles.

The combination of both questions produces a situation in which both the curricula of an integrated nature and the schemes that could be specific to integration are questionable. Therefore the concluding statement of the previous section reads: "What none of these schemes, nor any other existing scheme does, is allow us to evaluate, or even analyse, the methods or extent of integration in science curricular materials"; that is, as has been stated before, the schemes cannot consider integration more than the actual documents do if the categories of the analysis schemes have been constructed from the documentary on empirical evidence of the curricula claiming a unifying or integrating approach.

[1] P. Haussler and J. Pittman, A Curriculum Material Analysis System for Science, Weinhem and Basel: Beltz Verlag, 1973, p. 179.

The schemes can be, therefore, considered as analysis instruments for "normal" curricula and not specific to integration. An analysis system such as the CMAS of the IPN, Kiel, provides a basis for surface evaluation without telling much (if anything) about integration. If such a scheme were to analyse the integration aspects of an integrated curriculum, it would have to speak both to the content aspect and to the method aspect.

Questions as to the content aspect are: (a) what is integrated?, and (b) what are the principles used to integrate?

Question (a) contains a reference to the problem as to what can and should be integrated. Basically and traditionally, one of the main underlying concepts of science curriculum projects of an integrated nature is the assumption of a unified universe[1] and the basic unity of science[2].

The notion of unity has, of course, an older history. William James, for example, criticized in the last century the atomistic view point which attempts to

> "chop up the mind into distinct units of composition or function, numbering these off, and labelling them by technical names".[3]

Additionally, the view of a unitary whole, represented in the writings of John Dewey among others, supplements this concept development.

As for science, emphasis on the unity of the universe was made by the developments of the Wiener Kreis of the late Twenties - the neopositivistic circle around M. Schlick in Vienna, later represented by R. Carnap, H. Reichenbach and H. Feigl. Indeed, "unity of science" is the exact translation of Einheitswissenschaft, the term which was used synonymously for neopositivism: the journal of the Vienna Circle, Erkenntnis, continued as Journal of Unified Science after 1938.

Under the assumption of the "natural world in a piece", of course, nearly anything can be undertaken in order to restructure and reorganize science for instructional purpose. The unity pool thus provides the basis and elements for organizing principles which relate to the unified nature of science. The notion of unified science or the attempts to integrate attains thus the quality of operations within set theory where given elements can be organized according to sets of intended quality not, however, the not-yet given or recognized elements.

The problem now arising within the content area for integration concerns the validity of integrating within the sciences. Can one, therefore, legitimately talk of integration if science in its social context is not spoken to, if the linkage models do not cover such "external" disciplines as history or politics? The evaluator seeking a differential answer by means of an analysis scheme or a curriculum will not find himself in a position in which his needs will be satisfied.

The increasingly important field of science-related social issues and society-related science issues is not touched by any of the science education projects of an integrated nature, with very few possible exceptions such as the curriculum Man and the Environment, parts of the IPN Curriculum Physik and, for tertiary education, some modules of The Open University.

As therefore hardly any evidence exists for an integrated approach overlapping and extending the boundaries of science, establishing a truly multi-disciplinary content incorporating the controversial nature of science and society, the analysis schemes have not accounted for integration of this nature.

[1] Rutherford and Gardner, op. cit., 47 ff.

[2] V.M. Showalter, The Unity of Science. In: What is Unified Science Education? Columbus, Ohio: Centre for Unified Science Education, 1974, p. 11.

[3] L.T. Hopkins, Integration: Its Meaning and Application. New York: D. Appleton-Century Company, 1937.

Question (b) relates not only to the principles, but also to the value of these principles. What is indeed gained by reconstructing and reorganizing science content by operating from the "unified pool" only and seeing science education as a matter of set theory?

The function of a scheme should be to evaluate the value of the structuring principles of a science programme with an integrative nature. The value of a conceptual framework for cutting across the traditional divisions of the disciplines can then be seen to be increased if removal of the barriers enables science instruction to operate on the relevant levels mentioned above. Again, the function of the analysis scheme has to be seen as assessing the value of a structuring principle in view of stated criteria (e.g. in terms of relating its quality to a broader field of experience than provided by science alone).

From the method aspect, two main questions arise: (a) is and can a learning method be an integrating principle?, and (b) is there a method best suited to integration?

The method aspect shifts the problem from the object side (nature, science, technology) to the subject side (student, learning theory, cognition). It might even want to account for the fact that an interdisciplinary inquiry might be more the acceptance of the natural human process of exploration than a method, but nevertheless lift it to a method status for the operative purpose of construction of an integrated science curriculum.

This aspect is related to discussions in the philosophy of science concerning the logic of discovery versus the psychology of research, and linked with Dewey's theory of inquiry. The underlying assumption of both questions concerns conceptual understanding and is directed towards a method such as problem-solving or inquiry as a basis for integration. In this instance, the structuring principle of the subject matter is governed by the learning structure of the student. The analysis scheme should address: (a) the feasibility of methods in view of an integrated programme; (b) how to evaluate these methods with respect to their suitability to a content that takes into account the non-reduced scope of science - science as a process between man and nature by means of work; and (c) the integration of science disciplines with associated and inherent information such as the political, moral and economic implications.

When the content no longer distinguishes between "neutral" science and "value-oriented" non-science, when the point-of-view that "science is concerned with value-free statements" has made room for a more realistic approach, the method will have to account for these approaches. An example of the linkage and compatibility of method with content is given in the Humanities Curriculum Project of Schools Council/Nuffield Foundation.

The notion, for instance, of the "neutral chairmanship" in the humanities project (a rather basic teaching strategy being one of discussion rather than instruction) provides a basis on which the question "what is it about the method that is integrative?" attains a concrete status. Careful analysis of the content-method inter-relationship might lead to the statement that it is impossible to look at the aspect of integrating content without employing a method that is also integrative.

This has implications for teacher education, as the teacher's role introduces integrative features by means of chairing discussion groups which adopt the aim of the project. This issue of teacher training is naturally relevant also to the content side. Robinson has criticized the abstract use of "processes" or "methods" in science education and also stressed the importance of preparing teachers to overcome the distorted artificial structure in dichotomized science instruction by including philosophical considerations of science in teacher training, thus enabling teachers to avoid this dichotomy.[1]

[1] J.T. Robinson, Philosophy of Science: Implications for Teacher Education. Journal of Research in Science Teaching, Vol. 6, 1969, p. 99.

Finally, we should be compelled to analyse the role of evaluation in the development of the trends in integrated science curriculum construction. House has suggested that "there is a natural antipathy between change and evaluation" and that "evaluating something means being skeptical ...".[1] A penetrating consideration of these two statements could lead one to postulate that evaluation of integrated science curricula might negate the further development of endeavours in this area. This latter contention can, of course, be refuted if the results of the evaluation, although based in traditionality and skepticism, are positive.

The evaluator, and most importantly, the relationship with the curriculum developer (the client) come likewise into play in a consideration of the relevance of evaluation. Tawney (Chapter 2) discusses the difficulties of formal evaluation in relation to other groups who can be seen as clients of the evaluation. Contrary to the above possible interpretation, House contends that:

> "to induce change one must break through the image of the individual, the paradigms of the discipline, and the ideology of the organization".[2]

A person acting as an evaluator is one who dissents from doctrine, who seeks to introduce new information and to expose the curriculum developer to inconsistencies or contradictions in a programme. Perhaps in our thinking about the evaluation of integrated science curriculum materials, we need most of all to reflect upon securing evaluators who have a kindred but nevertheless questioning stance to development in this field.

[1] E.R. House, The Relevance of Evaluation. In: R.M. Rippey (ed), Studies in Transactional Evaluation. Berkeley: McCutchan Publishing Corporation, 1973, p. 256.

[2] ibid., p. 265.

ANNOTATED BIBLIOGRAPHY

Eraut, M., Goad, L. and Smith, G., The Analysis of Curriculum Materials. Brighton, Sussex: University of Sussex, Educational Area, Occasional Paper 2, 1975.

This publication presents a description of the background thinking which determined the structure of the Sussex scheme and a description of the uses of the scheme as well as the scheme itself.

Haussler, P. and Pittman, J., A Curriculum Material Analysis System for Science. Weinheim and Basel: Beltz Verlag, 1973.

The volume presents the rationale for development of an analysis scheme for science materials, a succinct discussion of the factors underlying the various categories of analysis, the scheme itself and a report on a study determining the inter-analyst reliability of the scheme.

Hopkins, L.T., Integration: Its Meaning and Application. New York: D. Appleton-Century Company, 1937.

Integration is discussed from psychological, pedagogical and sociological viewpoints. For Hopkins, integration is more an emphasis by the teacher than a special content or method.

House, E.R., The Relevance of Evaluation. In: R.M. Rippey (ed), Studies in Transactional Evaluation, Berkeley: McCutchan Publishing Corporation, 1973.

The author considers particularily the role of the evaluator in relation to the meaning of evaluation and its function in affecting change.

Robinson, J.T., Philosophy of Science: Implications for Teacher Education. Journal of Research in Science Teaching, Vol. 6, 1969.

The role of the teacher, in determining an instructional style, is reviewed particularily as it is seen as being largely dependent upon the teacher's background concerning the nature and philosophy of science.

Rutherford, J. and Gardner, M., Integrated Science Teaching. In: P.E. Richmond (ed), New Trends in Integrated Science Teaching, Paris: 1971, Vol. 1.

The authors present a rationale for integration and then consider how content can be selected for integrated science programmes, what methods are useful in the organization of such courses and seven practical considerations which must be taken into account when developing such courses.

Scriven, M., Value Claims in the Social Sciences. Boulder, Colorado: Social Science Education Consortium, Publication No. 123, 1966.

Although attending primarily to social science education, this article has much relevance to the aspects of science instruction related to values and value judgements.

Showalter, V., The Unity of Science. In: What is Unified Science Education? Columbus, Ohio: Centre for Unified Science Education, 1974.

This is a collection of articles describing various aspects of unified science education including a discussion of general consideration, terminology, the unity of science and guidelines for developing unified science units.

Smith, M.B., Editorial, J. abnorm. soc. Psychol., No. 63, 1961.

This short editorial, directed to research in psychology, points out nevertheless, the deficiencies in the guidance of scientific development by theory in light of the various historical perspectives.

Stake, R.E., Objectives, Priorities, and other Judgement Data. Review of Educational Research (Boulder, Colorado), Vol. 40, No. 2, 1970.

The author maintains that judgements are an integral part of evaluation. He reviews several sorts of judgement data and methods of gathering and reporting such data.

SSEC: Curriculum Materials Analysis System. Boulder, Colorado: Social Science Education Consortium, Publication No. 144, 1971.

This publication presents the Intermediate form of the SSED analysis scheme. A more comprehensive form (The Long Form) for professional evaluators and a shorter form (The Short Form) for teacher use are available from the same source.

Humanities Project. London: Heinemann Educational Books, Ltd., 1970.

This social science project for students in the age group 14-16, emphasizing the neutral chairmanship of the teacher in class discussion, aims at developing an understanding of social situations and the controversial issues that arise from such situations. The various units (8) and the forms of the materials should be of interest to those developing an integrated science curriculum.

IPN Curriculum Physik, Steureung und Automation (Regulation and Automation). Stuttgart: Ernst Klett Verlag, in press.

This unit is a good example of how non-science content can be integrated with science content. It is available only in Germany.

Man and the Environment. Boston: Houghton Mifflin Company, 1971.

An example of a one-year course using a content scheme as the integrating factory. Many of the influences on the environment (biological, physical, social, industrial) are incorporated.

The Open University. Bletchley, U.K.: The Open University Press, 1970.

The Science Foundation Course Units (34) developed for the Open University courses, involving studies in the various sciences, in the history of science and in the relation of science to society, illustrate how a unit approach might be used to create an integrated science curriculum.

FURTHER READING

1. The reader interested in further information concerning curriculum evaluation might well refer to the following:

AERA Monograph Series on Curriculum Evaluation: Vol. 1: Perspectives of Curriculum Evaluation, Vol. 2: Evaluation Activities of Curriculum Projects, Vol. 3: Instructional Objectives. Washington D.C.: AERA, 1967-9.

House, E.R. (ed): School Evaluation, The Politics and Process. Berkeley: McCutchan Publishing Corporation, 1973.

Bloch, J.R.: Quality and Quantity in Empirical Evaluation Studies. Studies in Educational Evaluation, Vol. 2, No. 2, 1976.

Popham, W.J. (ed): Evaluation in Education, Current Applications. Berkeley: McCutchan Publishing Corporation, 1974.

Schools Council Research Studies: Evaluation in Curriculum Development: Twelve Case Studies. London: MacMillan Education Ltd., 1973.

Walberg, H.J. (ed): Evaluation Educational Performance. Berkeley: McCutchan Publishing Corporation, 1974.

2. A more theoretical discussion of the possibilities and implications of integration of the sciences with non-science content can be found in the following:

Frey, K. in collaboration with Bayrhuber, H., Bloch, J.R., Dumpert, K., Haft, J., Hameyer, U., Haussler, P., Jaeckel, K., Kunzli, R.: The Concept Integrated Science Curriculum. Background paper to the Exeter Symposium on Integrated Science: and in: Studies in Science Education, 1976 (in print).

Frey, K.: Integrated Science Education in the Federal Republic of Germany. Presented at: Symposium on "An Interdisciplinary Approach Toward the Teaching of Science Subjects at the Secondary Level". Exeter, 8-14 September 1974. Strasbourg: Council of Europe, 1974 (mimeographed manuscript).

Fuller, W. (ed): The Social Impact of Modern Biology. London: Routhedge and Kegan Paul Ltd., 1971.

Habermas, J.: Theory and Practice. London: Heinemann Educational Books Ltd., 1974.

Habermas, J.: Knowledge and Human Interests. London: Heinemann Educational Books Ltd., 1972.

Lakatos, I. and Musgrave, A. (eds): Criticism and the Growth of Knowledge. London: Cambridge University Press, 1970.

Marcuse, H.: One-Dimensional Man. Boston: Beacon Press, 1964.

Margenau, H. (ed): Integrative Principles of Modern Thought. New York: Gordon and Breach, 1972.

Mischel, T. (ed): Cognitive Development and Epistemology. New York: Academic Press, 1971.

Phenix, P.H.: Realms of Meaning. New York: McGraw-Hill Book Company, 1964.

Rose, H. and Rose S.: Science and Society. Harmondsworth: Pelican Books Ltd., 1970.

5 What skills are needed to teach integrated science, and how can their development be monitored?

Clive Sutton
University of Leicester, United Kingdom

SUMMARY

This paper is mainly for those decision-makers who are in a position to assist the growth of a more skilful teaching community and who therefore need information about current levels of skill. These decision-makers include: (i) teachers individually and in groups, who are consciously trying to improve their own work; (ii) inspectors, advisers, and organizers of in-service study; and (iii) those reponsible for initial training.

It assumes that teacher skills are at least as important as curriculum materials and discusses ways of identifying those skills that are particularly required for an innovative programme of integrated science.

Trends in the demands made upon integrated science teachers are surveyed and suggestions are made for improving the flow of feedback to each of the above groups of people.

Why skills are of increasing interest.

The trend to discussion of teaching skills is partly the result of dissatisfaction with the supposedly 'teacher-proof' curriculum packages of the sixties. Members of curriculum project teams have sometimes expressed despair when their materials are used, but not in the way intended, and this has turned attention away from the books and apparatus, towards the patterns of interaction in the classroom. How far, for example, does the teacher encourage the speculative thought and systematic enquiry that the course designers intended?

There are at least two aspects to the question of skills. Firstly, supposing I am a teacher involved in a new programme of integrated science, am I myself skilled in these methods of enquiry? How good am I, for example, at suggesting hypotheses, designing experiments to test them, interpreting unfamiliar data, in an integrated science context? Secondly, do I possess the <u>pedagogic skills</u> to organize a learning environment in which children will conduct enquiries?

Some research suggests that even when teachers intend to teach in an enquiring style, they may unintentionally extinguish enquiry by the pupils instead of encouraging it.[1] The pedagogic skills seem to be of considerable subtlety. Will I and my colleagues acquire them without help, or are specific courses required? How much money and effort should be put into developing materials on the one hand and developing skills on the other? Certainly because teachers are

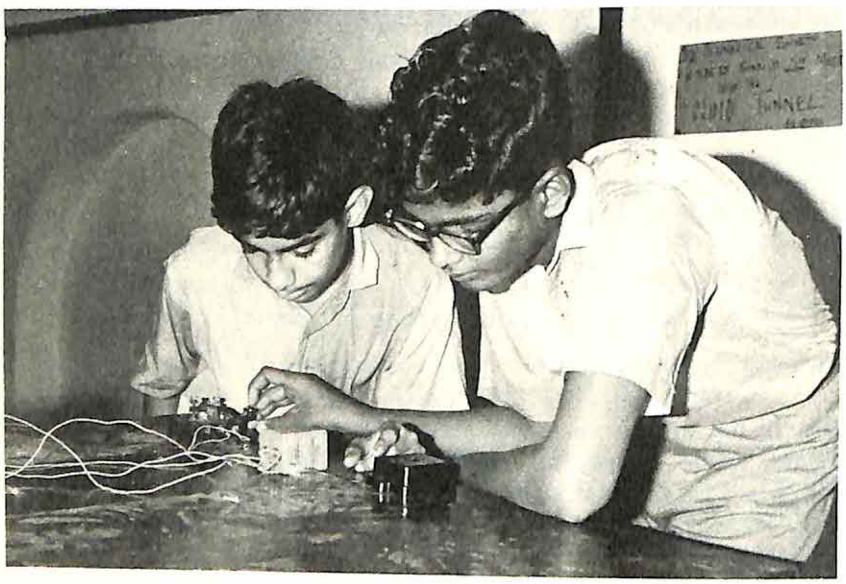
[1] See for example, D. Barnes et al., <u>Language, the Learner and the School</u>, Penguin, 1969, p. 25. (Tape recordings and group discussion as a way of developing understanding of teachers.), and M.R. Rowe, "Wait-time and rewards as instructional variables, their influence on language, logic and fate control. Part one: Wait-time". <u>Journal of Research on Science Teaching</u>, Vol. 11, No. 2, p. 81-94 (1974), also reported in Rowe, M.B., <u>Teaching Science as Continuous Enquiry</u>, McGraw-Hill (1973).

frequently viewed as managers, or facilitators of learning, these skills are an important focus of attention for evaluators. How and when are they acquired, and what factors assist or hinder teachers in the adoption of new roles?

What do we mean by skills?

The term "skill" implies something that can be acquired or improved. This contrasts with personality characteristics, like "intelligence" or "warmth", which so many researchers have tried (almost always unsuccessfully) to correlate with pupil achievement. We may extend the examples given in the previous section to include, say, "the ability to organize a room in which children will learn independently", or "listening skills", or "questioning skills", or "the ability to lead a discussion". These are by no means unambiguous in meaning, but they do suggest the possibility of practice and improvement. As we shall see below, some workers have preferred stricter operational definitions.

Indian boys working independently with electric circuits



To identify components of teaching is an analytic process intended to clarify one's thinking about what the job involves. It could be helpful both in diagnosing how to improve and in actually improving one's work. The analogy is with a musical performance - complex, variable, depending on the simultaneous use of many different sensitivities and skills that interact one with another, but capable of being improved by the practice of quite specific aural discriminations or manual dexterities. To describe and practise these does not exclude the possibility that some performers may achieve the same result by quite unorthodox techniques.

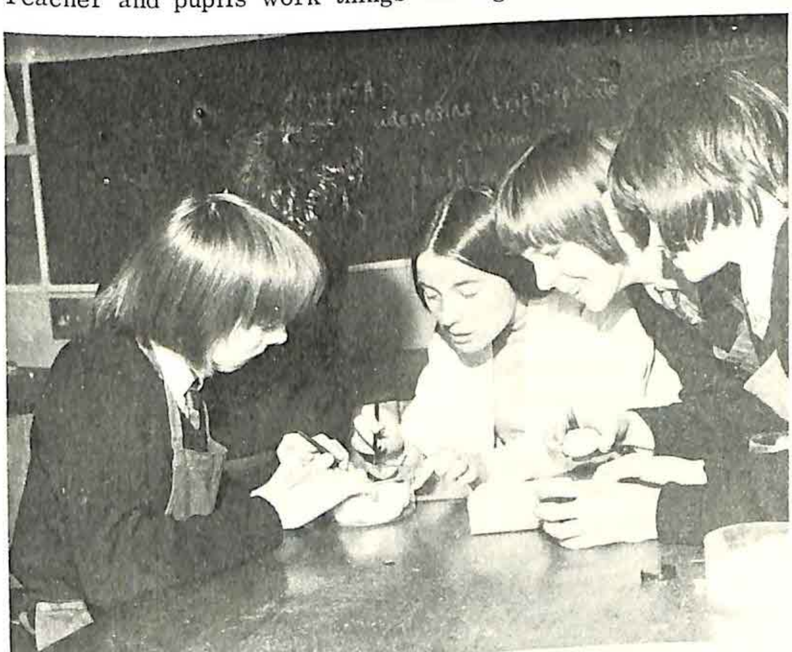
For example, Rowe describes a common form of questioning in American science lessons ("the inquisition") in which pauses by the teacher are of the order of one second in duration. The structuring of the argument is substantially done by the teacher, and pupils respond, at the teacher's bidding, in brief phrases, without any complexity of explanation. When teachers lengthen the pauses to about three seconds, particularly those which occur after a pupil has spoken, several changes occur. The pupils not only elaborate their first answers, but also make more comments and ask more questions on their own initiative. There are more speculative "suggestions" and more comparisons of one child's ideas with another's, and so on.

If we assume that such things are desirable at least for some of the time, the correlation is interesting, even though we might not be sure that it is the longer pauses which actually cause the greater pupil participation. It is a short step to saying that we have here an identifiable

skill which teachers could usefully practise. But at what level of generality should we define that skill? Shall we call it 'wait-time management' or 'ability to engage children in group discussion'? The former is more specific, and more amenable to simple practice and training; but on the other hand, it is possible to imagine teachers slavishly managing wait-time and getting no results, because they cannot simultaneously manage other factors such as the non-verbal means of engaging children's interest and building their confidence, the distribution of rewards in the group, or the provision of adequate stimulus. These are probably all subsumed in the 'ability to engage children in group discussion'. Perhaps we should tolerate the comparative vagueness of this latter term (e.g. just what quality of discussion is implied?) for the sake of the gain in reality.

If we take the analytic dissection too far, we get so many components that it would never be feasible to train anyone in all of them separately. We also risk asking teachers to practise things which they cannot afterwards put together into an effective teaching style. On the other hand, it could be that many components take care of themselves and that alteration of a few key ones could produce large changes in the total pattern of interaction. Rowe evidently thinks this is true for managing wait-time and for controlling the verbal rewards accorded to pupils. She has shown that teachers can change their behaviour in these areas by practice, with some of the results already mentioned. These changes might be establishment of different habits. Alternatively, they could be looked at as the external results of 'inner' changes in the teacher - changes in attitude, emotional changes (less fear of embarrassing silences), changes in how one sees the job. The way in which we look at them matters a great deal when it comes to devising programmes to improve teaching and learning, as I shall discuss in a later section.

Teacher and pupils work things out together



The notion of separate skills, and of 'teaching' as involving many of these skills in varying combinations, has considerable consequences for assessment and evaluation of teaching. For example, it makes possible a description of performance as a <u>profile</u> (rather than a global "good" or "bad"). It facilitates a move towards self-assessment of one's own work. It is possible that teachers fear assessment of performance as judgemental of themselves; but by discussing particular skills, we can bring to bear the non-judgemental stance of criterion-referenced testing and of mastery learning, in which it is a quality or skill which is assessed, not the person's over-all worth. For example, as a teacher, I may have mastered the skill of translating a

recall question into a comprehension question, but may never have attempted to lead a discussion on the social implications of scientific discoveries. As a decision-maker in relation to my own professional development, I will of course have to "take one consideration with another". If I lack a particular skill, I may take steps to acquire it (or in some circumstances I may defer doing so) and build upon others that I possess already.

Such evaluation is intended to be supportive of teachers in their own professional development rather than a demand for conformity to some predetermined pattern of teaching. There is no single definitive set of ideal teacher behaviours and it seems unlikely that there ever will be, given the great diversity of teachers, pupils, work situations and outcomes sought from a teacher's work. At best, the multi-skill view of teaching enables individual teachers to widen their repertoire of ways of working. The choice of which skills to use in any particular situation will depend on teacher judgement.

Trends in the demands made upon integrated science teachers.

Anyone reading the teachers' guides to recent integrated science courses is likely to be struck by the very great professional demands that are made upon the teachers.

Consider the following, from A Guide to the Australian Science Education Project (ASEP):

> "Teachers' views of their detailed responsibilities differ greatly as each interprets his or her broad responsibilities in ways consistent with his or her own strengths and the particular situation concerned ... As a basic to work from during the early development of ASEP, responsibilities of a teacher with respect to learning were categorized into four broad areas ..."
>
> "1. Providing facilities for learning.
>
> 2. Preparing the students for learning.
>
> 3. Effecting learning, and
>
> 4. evaluating the effectiveness of the learning that takes place and of the learning processes ..."[1]
>
> "The following is a summary of what (trials) teachers regard as being important when ASEP materials are being introduced and used ..."

There follow fifteen pages of comment on such things as knowledge of the broad aims of ASEP and their own curriculum goals, thorough preparation of the science room, organization of equipment, books and visual aids in the right place at the right time, handling individual and group work with different children doing different things at the same time, knowing when exposition is required in the form of formal teaching or demonstration, ability to keep track of pupils' progress, providing additional stimulus for the very bright and help and encouragement for the very slow, providing an atmosphere of trust, willingness to modify and improvise when necessary ("carry a teacher's kit of odds and ends in a bag or large pockets"), willingness to learn from the learners. The ASEP teacher is expected to establish rapport with pupils, help them to help themselves, move about the class and become involved in their activities, manage their safety, communicate, motivate, encourage and assess, guide pupils in the choice of activity, be enthusiastic, be prepared to listen, understand individual differences, and plan ahead. Specific mention is made of the skills of organizing discussions - in the arrangement of furniture, guiding pupils' preparation, monitoring the group dynamics of the event, and 'with tact, sensitivity and skill', redirecting the discussion if necessary, but without dominating it. An ASEP teacher will also have to be skilled in diagnosis beforehand and assessment of what is learned.

[1] Australian Science Education Project, A Guide to ASEP, Victorian Government Printer, Melbourne.

Elsewhere in A Guide to ASEP, it is clear that a teacher will need to know a good deal about children's intellectual development and about Piagetian theory and technique, and to have a critical awareness of the strengths and limitations of different types of enquiry. In addition to this, the teacher will need to be willing and able to find out about the major biological, chemical, physical, geological, sociological and psychological concepts that could be unfamiliar, and able to help pupils design scientific methods of investigation and analysis for problems that raise during their enquiries.

The ASEP materials are intended to be supportive to teachers; the aims of serving general education and pupils' personal development are ones which are widely acclaimed. The methodology of pupil activity, guided, but partly under their own control, while challenging to the confident teacher, may be daunting to others. ASEP is possibly more demanding than some other integrated science courses, but it is salutary to compare these expectations with those that have been made of some single-subject science teachers in the past. Roughly, they were to teach the content of the subject as they had already learned it, mainly by exposition, and train the pupils in the ability to answer examination questions about it.

The demands made by the Science 5/13 Project[1] upon primary and some secondary school teachers in the United Kingdom are similar or greater. Many of the primary teachers concerned are known to have very little background of science, and the booklets provide a lot of support, both in suggesting pupil activities and in providing background information for the teacher. Nevertheless, the flexibility implicit in the approach requires that a teacher should be able to notice what captures the interests of particular pupils, to discern the teaching potential in these interests, guide the children to concentrate on the more investigable questions, and manage their enquiries so that useful learning occurs in one or more of eight general categories of objectives. Again this is to be done for groups of children working on different topics in the same room.

The demands of some other integrated science programmes seem relatively modest, as at least the topics to be taught have been chosen by someone else. For the Scottish Integrated Science Scheme, the working party recognized that teachers were qualified frequently in only one or two separate sciences, and that they might nevertheless "be required to teach subject matter for which they have no great interest or aptitude". "We recommend throughout this report ... the teaching for attributes other than knowledge, and we know that we are asking many teachers to undertake work for which they may have little specific training". Corresponding problems have been met in the West Indies Integrated Science Project and Malaysian Integrated Science, both of which are adaptations of the Scottish scheme.

In sponsoring the course "Science: A Process Approach", the American Association for the Advancement of Science found it necessary to devise a full "Commentary for Teachers" and a "Guide for In-Service Instruction" in order that teachers might become skilled in "Basic processes" such as observing, classifying, measuring, predicting and inferring, and "integrated processes" such as controlling variables, formulating hypotheses, and experimenting. It was not assumed that teachers would have these skills by virtue of having studied science themselves!

For the Intermediate Science Curriculum Study[2] in the United States, the organizers have devised a series of individualized teacher preparation modules, several of which provide training in pedagogic skills, such as how to organize a classroom, how to use questions well, and how to monitor pupil progress.

The trend to demand of teachers more substantial understanding of educational theory is well illustrated in the Schools Council Integrated Science Project (SCISP), with its 'pattern-seeking approach'. Like the ASEP teacher, the teacher who uses SCISP will also discuss controversial

[1] Science 5/13, With Objectives in Mind, Macdonald Educational (1973).

[2] Intermediate Science Curriculum Study Individualised Teacher Preparation. Several volumes published by Silver Burdett General Learning Corporation, Morristown, New Jersey (1972).

issues in the classroom and take up a role which was formerly unfamiliar to science teachers - that of dealing with questions for which there is not one generally accepted answer. For some teachers, discussions of science and its social implications is more than they thought they were employed for.

A comprehensive set of materials for the development of science teacher skills is that prepared by the Science Teacher Education Project[1] in the United Kingdom. This bank of resource materials and ideas is not linked to any particular science course, either single subject or integrated. The project did not publish an exhaustive list of skills needed to teach science. However, the suggested activities are all designed to support the autonomy of the individual teacher, to encourage flexibility and develop powers of judgement and decision: e.g. "a willingness to look for the teaching potential offered by pupils' questions" (STEP Unit ADA/7).

Table 6. Summary of kinds of demand being made upon teachers of integrated science.

1. KNOWLEDGE from several sciences, and from educational theory*.

2. SKILLS OF ENQUIRY - both in information retrieval (finding out what they do not already know from books, etc.) and a range of 'science process' skills.

3. PEDAGOGIC SKILLS. Some of these are interpersonal - ways of relating with children; some of them are in management (e.g. arrangement of the learning environment, keeping records of pupils' progress, devising tests for this purpose); and some are curriculum planning skills (e.g. ability to see the teaching potential in the neighbourhood of the school, selecting feasible objectives and learning experiences).

* Knowledge and skills are not totally separate. Science concepts may be necessary for exercising science process skills. Educational theory may be required, to provide a conceptual apparatus that is indispensible to the development of pedagogic skills.

Some writers (e.g. Showalter) describe values and attitudes desirable for teachers of integrated science. An attitude is a tendency to perceive things in certain ways and to act in certain ways, so attitudes may follow from the possession of knowledge. Newly-acquired concepts effectively provide a new "set of spectacles" through which the world is viewed and this is the basis of the inter-relationship between cognitive and affective aspects of learning.

The introduction of integrated science courses also appears to have implications for the roles of teachers. A "single subject" teacher could find a personal identity partly by thinking of herself/himself as a 'chemist', or a 'physicist', an expert in the mysteries of a particular subject. An "integrated science" teacher appears by contrast to be more of a teacher and less of a subject person, teaching not in order to transmit the subject but in order to develop the pupils. The science is justified for its contribution to general education and not for itself alone. The "integrated science" teacher is expected to bring out connexions not only between different branches of science but also between school science and everyday life. Such a teacher is a guide in enquiry into the local environment and an interpreter of everyday phenomena. Bernstein[2] has described the pressures that this places on both teacher and pupil. The teacher is required to reveal more of herself/himself and to expect the pupil to do so as well, as cross-references are

[1] Science Teacher Education Project, Activities and Experiences, The Art of the Science Teacher, and Theory into Practice, McGraw-Hill, U.K. (1974).

[2] B. Bernstein, "On the classification and framing of educational knowledge", in Class Codes and Control (1973). (Related to a discussion of constraints on role identification for integrated science teachers.)

made continually between the lesson content and one's personal experience. Bernstein argues that the psychological safety accorded by the old curriculum, in which subjects are isolated from each other and from real life, is part of the reason for the survival of such a curriculum. The integrated curriculum, he argues, gives less privacy to teacher and pupil, and requires a more explicit consensus of values.

Attempts to develop and monitor skills.

Rather less work has been done on the development of science process skills than on the pedagogic skills. In some countries, there has been a growth of personal project work in initial training courses. In the United Kingdom in the sixth form and at college, students may nowadays be put in a position where they have to select for themselves a general area of enquiry, define an investigatable problem within it, state hypotheses with predictions, design experiments and carry them out, and interpret the results. Assessment schedules exist which discriminate a person's competence in these skills separately (see bibliography). Procedures more characteristic of one branch of science than another (e.g. design a control experiment in biology) no doubt require experience at least within these different branches.

By contrast to the lack of growth of process skills, there has been a huge growth in techniques for developing pedagogic skills. These techniques include (i) interaction analysis, and (ii) microteaching, which have built-in means of monitoring the acquisition of skills. Both feature frequently in the movement on the United States for (iii) competency-based teacher education.

(i) Interaction analysis as a training tool has been associated with the names of Flanders, Bellack and others. Some authors claim that very extensive training (of the order of thirty hours or longer) is needed before observer awareness of the classroom is changed substantially. Others favour alerting student teachers to 'things to look for' by means of observation record schedules used for much shorter times. Either way, the first target is the students' awareness of patterns of interaction and of alternative possibilities. Students may later practise adopting a particular style, in which case the observation schedule can be used as an evaluation instrument to show them how they are progressing. Most users of interaction analysis claim to avoid specifying preferred interaction patterns, but frequently imply that some are "better" than others.

(ii) Microteaching, though regarded by many as simply a means of practice for beginners whereby they can teach a few children for a short time, is properly characterized by two other features, namely immediate feedback (e.g. by replay of a video-recording), and concentration on a particular skill.

Examples of such skills are: varying the stimulus, establishing a mental set, drawing a lesson to a close, reinforcement skills, use of examples and encouraging pupil participation. In preparation for integrated science work, one might add items such as 'selecting examples from many different fields'. A microteaching session may be preceded by a seminar and by guided reading about the particular skill to be practised. Then the student teacher conducts a mini-lesson with a few pupils (say, five) for a short time (say, ten minutes). In this, he or she makes a deliberate effort to practise the chosen skill. A video-recording may then be played back during a discussion with the supervisor. The teacher may go on to a further session of mini-teaching. Many texts now exist illustrating the method (e.g. Brown). The costs of hardware are coming down in Western countries relative to the rising costs of tutorial visits to schools, and this may assist the growth of the technique. However, the substantial consumption of time and the organizational problems of providing groups of pupils readily available in the college tend to limit student teachers to practizing only a very few skills.

There are at least two approaches to microteaching. One is a strict behaviourist approach, implying that the 'skills' are involuntary habits and that old ones can be eliminated and new ones established by this sort of practice, accompanied by appropriate regard from the tutor. As mentioned, this could only be done for one or two skills per student teacher. The alternative

view is that, as with interaction analysis, the method "opens the eyes" of participants to possibilities they had not thought of and prepares them to observe other teachers and themselves more closely and with more perception.

(iii) Competency-based teacher education (CBTE) focusses on what it is hoped the teacher will be able to do (not just what it is hoped he will know). In 1968, the United States Office of Education invited specifications for programmes of teacher education of this type, and nine institutions subsequently had their proposals funded. Factors which contributed to the development of CBTE programmes probably included the emphasis on behavioural objectives and systems analysis in contemporary literature, disenchantment with existing programmes, and particularly educational theory courses, discussion of cost effectiveness and pressures for teacher accountability. "Competencies" (skills, knowledge, behaviours and sometimes attitudes) which it is hoped teachers will develop are derived by the CBTE designers from many sources. They may work from a job analysis, from "expert opinion" on what teachers will be required to do, from indications in empirical studies that particular skills may be linked with particular pupil gains, or else they may proceed "logically" (from philosophical assumptions about the nature and purpose of teaching and learning) to specify what competencies should be fostered.

Those concerned with integrated science courses which have a distinctive philosophical base (e.g. Science Education Program for Africa) might find this approach useful. However, there are a number of cautions to be made. The possession of a list of competencies appears to simplify the evaluator's task, since at least the desired outcomes are declared. On the other hand, it could draw attention away from unexpected and unintended outcomes, and from the process by which student teachers are learning. In my view, it certainly oversimplifies the complexity of meaning of "effectiveness" of teaching.

Table 7, over page, presents some arguments for and against CBTE programmes, in such a way as to raise questions about what evaluators concerned with new programmes in integrated science might try to monitor, in addition to the specified competencies.

Table 7. Some arguments for and against CBTE programmes, and corresponding considerations for evaluators.

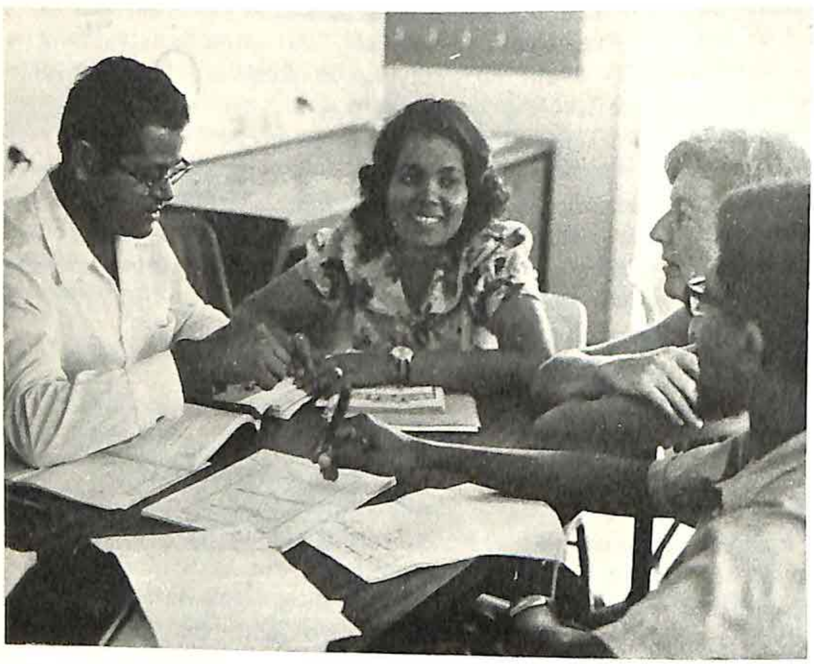
For	Against	Notes for evaluators
1. By identifying skills which an intending teacher may practise and master, CBTE allows beginners to concentrate on one thing at a time, instead of facing them with the immense complexity of the whole job. They may gain confidence rapidly.	Impractically long lists of microskills may have the opposite effect. Also, to maintain confidence the beginner will have to know that he can use particular skills in the more complex context of whole teaching and not just in isolation.	Evaluators need ways of monitoring skills not only separately but in the context of whole teaching of the innovative programme of integrated science. They might also attempt to follow teachers through initial and in-service courses, examining their motivation and response to the various learning tasks.
2. The emphasis on student performance assists the tutor out of the role of information-giver and into the role of a coach who helps the students to practise and develop their own competence.	This emphasis may encourage the belief that teacher preparation is just the acquisition of a set of behaviours, ignoring the likelihood that such behaviours are related to 'internal' qualities - attitudes, concepts, ways of looking at the job. It could result in laborious and ineffective attempts to acquire behaviours which would follow easily from changes in the student's cognition or affect.	Young teachers' perceptions of the job, and particularly changes in this in response to different experiences, are an important area of enquiry. When a change to an integrated science programme is made, changes in their view of their own role should be examined (see Combs).
3. The emphasis also on what it is hoped <u>pupils</u> will be able to do assists the student teacher out of the role of information-giver and into the role of helper or manager of children's learning.	Close specification of any aspect of the job may diminish the teacher as a professional, responsible for decisions about ends and means, and tend to make him or her into a technician applying specified procedures to produce given results.	How teachers make decisions about their work is possibly as important a topic of enquiry as what skills they have. Attention should be given to identifying the area of freedom that an integrated science teacher (a) has, (b) thinks he has.
4. Clarity of description should help both tutors and teachers to demonstrate accountability to themselves as professionals.	There is some political danger in drafting lists of competencies which become the focus of power struggles between say administrators and teachers, or parents and teachers. Within a CBTE framework of thought, it may be somewhat easier to dictate what teachers or institutions shall do.	A full understanding of any innovative programme will require a description of the social and political context in which it is occurring.

Suggestions for action.

Certain skills of enquiry and associated concepts used in an integrated science context may not have been developed in traditional single subject courses which the teachers experienced in their own early education. Almost certainly, also, some requisite pedagogic skills will not have been acquired during their training. Hence:

1. Teams of developers, in conjunction with teachers in a particular region, might usefully compile a statement of what skills are thought necessary to put the integrated science programme into effect. Lists of such skills should form a guide to be borne in mind by teacher educators and others. There is some value in attempting to state these in terms of what teachers are expected to be able to do, or at least to illustrate the desired skills by giving examples of what teachers might do. However, the possession of particular skills should not be mandatory upon teachers, as there is, as yet, no evidence of the indispensibility of any one skill.

Trinidadian teachers involved in curriculum development



2. Programmes to help teachers develop and improve their skills should not be launched without some consideration of the assumptions underlying the learning methods that are proposed. For example:

   (i) It might be thought that relevant skills can be acquired simply by practising them until they become established as habits (behaviour modification).

   (ii) An alternative view is that the behaviour of a teacher changes only in response to changes in the concept of the job and its purposes. (A viewpoint elaborated by Combs[1].)

[1] Arthur W. Combs, The Professional Education of Teachers; Allyn and Bacon Inc., Boston (1965).

3. There are several techniques intended to increase understanding of or to give practice in particular skills and which also provide a means of evaluating progress. They include: recording one's own lessons for subsequent discussion in groups; discussion of lesson transcripts; simulation and role play; interaction analysis; and microteaching. The cheapest and simplest of these is probably discussion of lesson transcripts and a brief description of such a method for an in-service group of teachers is given by Barnes. It was developed with groups of serving teachers, who discussed tape recordings and transcripts of lessons. In the process, they became able to attend to aspects of teachers' and pupils' language which had previously been unnoticed. Another programme is described by Adelman and Elliott[1], also working with serving teachers. These teachers were asked explicitly to consider whether in their classrooms there was evidence about a number of suggested hypotheses linking particular teaching skills with the children's readiness to enquire. In both cases, the teachers concerned seem to have acquired a new way of thinking about and looking at their work, i.e. a change in understanding came first, and the writers seem to think that this prepares the way for the development of new skills.

   The interplay of these two aspects is of particular interest. The increasing attention paid to "skills" in recent years had been at the expense of considering the relationship between knowledge and skills. There has been a tendency to think that the skills can be acquired in themselves, perhaps in a desire to get away from 'old-fashioned' teaching of large amounts of information and concepts. Teacher educators are talking more about what people can do and are less keen to be seen worrying about what they know. The question of whether <u>doing</u> can occur without <u>knowing</u> has meanwhile been avoided. Teachers can obtain feedback about their own skills not only by these techniques but also by careful monitoring of what their pupils can do, using the methods developed by Harlen. (See also Duckworth's checklist in Harlen's paper in this volume.)

4. Attempts should be made to ensure a flow of feedback to all relevant decision-makers, in a form that is not judgemental of individual teachers.

5. Consideration should also be given by curriculum planners to the constraints limiting the development of relevant skills. These might include the teacher's self-concept and educational history, the authority pattern of the culture and pupil expectations of teacher behaviour. Attempts to help teachers develop skills should take account to these constraints. For example, the development of integrated science teaching in a community where separate subject teaching is already well established, and where the teachers have themselves as learners never experienced an integrated study, may be extremely difficult. It might require the introduction of some integrated science courses in the universities and colleges, and the preferential selection of teachers from these.

[1] C. Adelman and J. Elliott, "Teacher Education for Curriculum Reform", An interim report on the work of the Ford Teaching Project, <u>British Journal of Teacher Education</u>, Vol. 1, No. 1, January 1975.

BIBLIOGRAPHY

1. Science Process Skills. A useful guide to the construction of in-service courses was published by the American Association for the Advancement of Science (AAAS), Guide for In-Service Instruction (1967), together with Science: A Process Approach, Commentary for teachers. L.E. Klopfer provided an interesting taxonomy of objectives reflecting science processes - see: Bloom, B.S., Hastings, J.T., and Madaus, G.F., Handbook of Formative and Summative Evaluation, McGraw-Hill (1971).

   The Federation for Unified Science Education has published in Prism II, Vol. 2, Nos. 3 and 4 (1974) a valuable discussion of the "dimensions of scientific literacy".

   Where it is hoped to develop teacher confidence and competence in enquiry by means of personal project, a useful analysis of what is involved in setting up and completing a scientific project is given in Nuffield Advanced Science Projects in Biological Science, Penguin (1970).

2. Interaction Analysis: For example, see Wragg, E.C., Teaching Teaching, David and Charles London (1974). For simpler methods, See Haysom, J.T., and Sutton C.R., (eds) Theory into Practice, McGraw-Hill (1974). This latter volume incorporates a wide-ranging set of observation record sheets designed to focus the student teacher's attention on one aspect of teaching at a time. They can be used in college, with filmed or videotaped lesson episodes, or in school.

   Eggleston, J.F., Galton, M., and Jones, M., A Science Teaching Observation Schedule, Schools Council (1975). This schedule, designed for research purposes, not for teacher training, nevertheless contains an interesting categorisation particularly of teachers' questions.

3. Microteaching. Allen, D., and Ryan, K., Microteaching, Addison Wesley (1969): Jensen, R.M., Microteaching: Planning and Implementing a Competency-Based Program, Charles Thomas, Springfield, Illinois (1974). Gives samples of micro-lessons. Brown, G., Microteaching, Methuen (1973).

4. Competency-based Teacher Education (CBTE). A full discussion can be found in a series of articles in Volume XXIV, No. 3, of the Journal of Teacher Education, Fall 1973. Research into the empirical basis for the selection of competencies is further analysed by Heath, R.W., and Nielson, M.A., in "The Research Basis for Performance-Based Teacher Education", Review of Educational Research, 44, No. 4, p. 463-484 (1974). They claim there is no empirical basis for the prescription of teacher training objectives because of poor operational definitions of both teacher skills and pupil achievement, and because of weak research design.

6 Evaluation instruments for integrated science teaching

Victor J. Mayer and James M. Richmond
Faculty of Science and Mathematics Education, Columbus, United States

SUMMARY

This chapter briefly reviews data-collecting procedures useful in evaluation studies of integrated science teaching. It focuses especially upon objective respondent-completed instruments identified in a world-wide survey of science educators. The instruments are categorized and representative examples of each category are discussed. Recommendations are suggested for instrument use and development.

Various data-collecting procedures have been developed for use in evaluation studies that are adaptable for use in integrated science. The decision as to which procedures are to be used in a given study will depend upon several factors, including the nature and use of the information desired, the size of the implementation or development project, the approach to be used in evaluation and the resources available to the evaluator. Procedures that have been extensively used include interviews, classroom observation, free-response forms, projective techniques, questionnaires, tests and checklists.

The more informal procedures have the advantage of providing a wide variety of detailed information. Interviews, for example, allow the evaluator to probe teachers for potential problems implicit in the curriculum materials being used and to follow-up unanticipated characteristics of the implementation effort as raised by the teacher in the interview. Such informal means, therefore, are most useful in formative stages of evaluation, when the developers themselves are likely to be conducting the evaluation and to be using the resulting information directly in the revision of materials. Such intensive data-collecting techniques, however, require much staff time and would be extremely expensive to utilize on a large scale, especially in the summative phases of evaluation. They also yield information that may be suspect, because of possible bias on the part of the data collector or as a result of egocentrism on the part of respondents to interviews and open response forms. In addition, the presence of an outside element (the data collector) in the classroom, or the interaction of strangers with the teacher or child may alter the context to such an extent that the information obtained is not representative of the actual climate in which the curriculum is being used.

In this paper, the detailed examination of data-collecting procedures has been limited to respondent-completed paper-and-pencil forms that are capable of yielding a single score on one or more separate factors. They are what are referred to in this paper as "instruments"[1].

[1] The reader interested in "classroom observation systems" is referred to the comprehensive review by A.L. Balzer, T.P. Evans and P.E. Blosser, *A Review of Research on Teacher Behaviour Relating to Science Education*, Association for the Education of Teachers of Science and the ERIC Science, Mathematics and Environmental Education Clearinghouse, Columbus, Ohio, 1973, p. 396-447.

The data-gathering instruments discussed herein cannot provide as rich and detailed information as the more informal techniques. But they are less expensive to use in terms of staff time, and therefore can be used with a larger population to yield a wider base of information. Also, by using techniques such as matrix sampling, the amount of information obtainable from each administration of such instruments can be increased. Further, by using properly developed and administered instruments, objective and representative information can be obtained through the elimination of observer bias and by the removal of outside influences that might be introduced into the classroom by use of informal data-gathering techniques.

The purpose of this paper is to provide the reader with an understanding of the wide variety of evaluative instruments that have been developed and the broad range of information which can be obtained through their use. To this end, the types of instruments available have been categorized and examples selected for description from each category.

An additional function is to identify instruments which can be used as models for the development of other instruments for use in evaluation studies. In a few cases the actual instrument itself might be useful. Information is provided in Table 8 and in the list of references that will allow the reader to locate existing instruments.

The instruments discussed were selected from several sources: Unpublished Evaluation Instruments in Science Education: A Handbook[1], which describes 293 instruments; responses to a worldwide questionnaire survey of science educators conducted in June 1975 which yielded an additional thirty-five instruments; and from a review of science education literature since the publishing of the Handbook. Instruments were selected according to the following criteria, namely:

1. Availability, i.e. the instrument is available from one of several sources - ERIC microfiche collections, a journal article, directly from the author or his institution or - where the source is a dissertation - through University Microfilms.

2. General Usefulness, i.e. many instruments are available that have been designed to evaluate local curricula or to determine cognitive or affective outcomes of specific instructional media. Unless such instruments presented a novel format or design, they were not selected.

3. Usefulness in evaluation integrated science curricula. Most of the instruments selected are listed in Table 8, over page, along with selected characteristics, author and references to the bibliography.

[1] Victor J. Mayer, Unpublished Evaluation Instruments in Science Education: A Handbook, Center for Science and Mathematics Education, The Ohio State University, Columbus, Ohio, 1974.

Table 8. List of selected instruments.

Title	Items	Reliability	Validity	Author
Student Characteristics				
Understanding in Science Test	See text	See text	Concurrent	Tisher and Dale
Science Reading Test	Multiple-choice	0.90-0.92	Content	Heath
Mathematics Skills Test	Multiple-choice	0.96	Content Concurrent	Denny
Scientific Curiosity Inventory	Statement sets	0.85	Content	Campbell
Scientific Reasoning Test	Multiple-choice	0.67	Content	Lucas
Critical Thinking Battery	Multiple-choice	0.90	Content Concurrent	Cillizza
Home Attitude Instrument	Likert-type	-	-	Starring
Science Experience Inventory	Statements	0.93	Content	Ulhorn
Student Performance				
Test of the Understanding of Science	Multiple-choice	0.76	Content	Cooley and Klopfer
Science Process Inventory	Statements	0.86	Content	Welch
Test of Social Aspects of Science	Likert-type	0.71	Content	Korth
Abridged Scientific Literacy Instrument	Likert-type	0.76	Content	Richardson and Showalter
Pictorial-Aural Inventory of Science Knowledge	See text	0.73	Content	Finkelstein
Level of Understanding of Science Concepts	Multiple-choice	0.90	Content	Helgeson
Concept-Process Test	Multiple-choice	0.83	Content	Disinger
Fyffe Process Test	Multiple-choice	-	Concurrent	Fyffe
Test of Science Processes	Multiple-choice	0.91	Content	Tannenbaum
Basic Science Processes Test	Multiple-choice	0.35-0.71	Content	Beard
Test on Scientific Attitudes	Multiple-choice	0.55	Construct	Nay and Kozlow
Cognitive Preference Examination II	Multiple-choice	0.70-0.77	Content	Atwood and Wright
Beliefs About and Attitudes Toward Science and Scientists Scale	Multiple-choice	0.57	Content	Champlin
Inventory of Societal Issues	Likert-type	0.62-0.77	Construct	Steiner
A Test of Interests	Likert-type	0.75	Construct	Meyer
Attitudes to Science	Likert-type	0.69-0.91	Content	Brown
Climate or Process				
Science Classroom Activity Checklist	Statements	0.77	Content	Sagness
Science Classroom Inventory	Statements	0.82	Construct	Shay
Earth Science Test	Multiple-choice	0.88	Content	Mayer, et al
Test of Knowledge of SAPA	Multiple-choice	0.68	Content	Merkle
Science Support Scale	Likert-type	0.84	Content	Schwirian
Rating Scale of Objectives of Science Education	Statements	0.93	Content	Berty

DESCRIPTIONS OF INSTRUMENTS

In this section, examples of instruments will be discussed in three categories based upon the broad types of variables they purport to measure: student background characteristics; student performance variables; and climate or process variables.

The first group of instruments consists of those that attempt to assess student background characteristics that are probably stable for the purposes of a particular investigation; e.g. stage of intellectual development. They form the student context in which a curriculum is being implemented or tested. Knowledge of such variables is important in both formative and summative evaluation stages in order to specify which student characteristics influence the degree of success or failure of the curriculum implementation effort and to provide the developer with guidance in constructing materials and instructional strategies appropriate for different kinds of students.

The second group of instruments measures the output variables of a curriculum; the types of student performance the curriculum is designed to elicit, such as science knowledge and attitudes toward science. The instruments reviewed in this category may be useful in summative evaluation stages, where there may be an attempt to differentiate between the effects of two or more curricula. In the formative evaluation stages, the instruments may be useful as models for instrument development.

The third group measures process or climate variables: the characteristics of the classroom environment in which the curriculum is being implemented. This category of instruments is useful in summative evaluation and can also provide models of instruments to be developed for use in formative evaluation, to provide control over the degree to which the classroom environment supports the objectives of the curriculum being evaluated during its implementation.

Instruments that assess student background characteristics

This first group of instruments relates to those student characteristics concerned with a student's ability to learn from science instruction and materials; characteristics such as scientific reasoning and critical thinking, which may in turn be affected by science instruction but probably not appreciably in the time framework of the typical science course. These characteristics, although they may fall within the domain of objectives suggested for science instruction, are of a broad nature and if affected by instruction, they are affected by the total school context and/or are contributed to incrementally by instruction over a long time-frame (years, rather than semesters). Therefore, for the purposes of most evaluation efforts in integrated science, they can be thought of as relatively invariant characteristics of the student.

(a) Level of cognitive development

The Australian Science Education Project (ASEP) materials have been designed for use at three levels of student cognitive development. Stage 1 materials relate to Piaget's concrete level of development, Stage 2 are suitable for students in the transitional period between concrete and formal operations and Stage 3 materials deal with hypothetical situations and are designed for students at Piaget's formal operations level. Tisher and Dale have developed a pencil-and-paper instrument called Understanding in Science Test[1] which may prove to be valuable to the science teacher as a quick and efficient method for determining the level of cognitive development of the students in his classroom. The test, which takes about forty minutes to administer, provides a measure of the student's understanding of physical science concepts in four broad conceptual areas, namely, reflection from a plane surface, the principle of the balance, balancing volumes of liquids, and the projection of shadows. Before responding to the test items, the students are either provided with a demonstration of the four sets of apparatus in question, or shown diagrams of the apparatus by means of an overhead projector. The Understanding of Science Test was validated with a sample of fifty-seven Australian students in

[1] R.P. Tisher and L.G. Dale, "Understanding in Science Test", Australian Council for Educational Research, Frederick Street, Hawthorn, Victoria, 1975.

grades 7, 8 and 9, ranging in age from 12.0 to 16.4 years. Since the instrument was developed as a pencil-and-paper alternative to the clinical interview, all subjects who took the test were also interviewed on "invisible magnetism", "equilibrium in the balance", and "combination of chemicals" as described by Inhelder and Piaget. The authors reported a close agreement between the two methods of cognitive classification; however, the pencil-and-paper test tended to classify the students slightly higher than did the clinical interview.

(b) Reading and mathematical ability in science

Several types of instruments have been developed to assess student reading or mathematical abilities related to science. Examples of such instruments are the Science Reading Test[1] (which assesses achievement of selected reading skills as applied to science) and the Mathematics Skills Test[2] (developed to measure student performance at three cognitive levels on ten basic mathematics skills related to chemistry learning). The mathematical skills are computation, signed numbers, use of parentheses, fractions, decimals, exponents, percents, one-variable equations, ratio and proportions, and graphing. Both were developed specifically for use in American school situations.

(c) Intellectual characteristics

A certain degree of confusion seems to exist in the literature concerning the identification and classification of intellectual characteristics of students that affect science learning. The following three instruments, however, offer ways of measuring what would seem to be somewhat distinct characteristics:

(i) The Scientific Curiosity Inventory[3] assesses the degree of curiosity that junior high school students express concerning various aspects of science. It consists of seven sets of statements. Within each set the respondent is asked to answer yes or no to each statement in the context of two questions which are posed at the beginning of the set.

(ii) The Scientific Reasoning Test[4] purports to assess the respondent's ability to judge the validity of evidence, to design experiments and to comprehend scientific writing. It consists of a series of situations in which the respondent is given written descriptions and/or data concerning experiments or observations. The respondent then answers a series of multiple-choice items based upon the situation.

(iii) Critical Thinking Battery[5], developed by Cillizza, has four sections, history, English, science, and mathematics, each consisting of sixty multiple-choice items, subdivided into three parts corresponding to critical thinking skills in the particular content area. Each of the three instruments was developed for use with children in English-speaking countries; the United States and Australia.

(d) Home background and past experiences

The lack of attention paid by evaluators to the effects of student home environment upon their ability to profit from a particular curriculum is evidenced by the paucity of instruments developed

[1] Mayer, Unpublished Evaluation ... op. cit., p. 110.

[2] ibid., p. 38.

[3] ibid., p. 107.

[4] Arthur M. Lucas, "Scientific Reasoning Test", School of Education, Flinders University of South Australia, Bedford Park, 1969.

[5] Mayer, Unpublished Evaluation ... op. cit., p. 104.

in this area. The Home Attitude Instrument[1] is the only one found which is designed to assess the influence of the home environment upon the attitudes of students toward science and scientists. The Science Experience Inventory[2] was designed to determine the types of science-related experiences that were a part of the student's background. Both were developed for use in American schools.

Instruments that assess student performance

In this section, instruments are reviewed which purport to measure those aspects of student performance which are generally thought to be amenable to change through relatively short-term science instruction. These include instruments assessing attitudes toward science and scientists, knowledge of science and ability to use science processes. This is where the greatest developmental activity has occurred, particularly over the past ten years, when there has been increasing cognizance of the function of science curriculum in the development of attitudes toward science and in the development of the ability to use problem-solving procedures. The evaluation of nationally-supported curriculum development efforts, such as those of ASEP and of the variety supported by the National Science Foundation in the United States, seem to have been largely responsible for this flurry of activity in instrument development.

(a) Understanding the nature of science

A number of instruments have been developed over the past fifteen years which attempt to assess a respondent's understanding of various aspects of the scientific enterprise. The one that has been most widely used on a worldwide basis is that published by Cooley and Klopfer in 1961, entitled the Test of the Understanding of Science[3]. The instrument is of three parts, assessing respectively the understanding of the scientific enterprise, the characteristics of scientists and the methods and aims of science. An instrument with similar objectives, the Science Process Inventory[4], was developed for use with American High School students and adults. The respondent is asked to agree or disagree with each of 135 statements about science and scientists. Korth[5] developed an instrument which focuses specifically on the interaction of science and society, for use with American students. The Abridged Scientific Literacy Instrument[6] was developed to assess attitudes toward science and understandings of inter-relationships in science. The instrument consists of thirty-four situation-establishing items, each with a seven-point Likert-type-response scale. It was used with American high-school graduates enrolled as college freshmen.

(b) Achievement in science knowledge

A wide variety of instruments, many of which have been published commercially, has been developed to assess student levels of knowledge in various science disciplines and in science generally. Those published in the United States are reviewed in a publication by Wall and

[1] ibid., p. 99.

[2] id.

[3] Janet Wall and Lee Summerlin, Standardized Science Tests: A Descriptive Listing. National Science Teachers Association, Washington, D.C., 1973.

[4] Mayer, Unpublished Evaluation ... op. cit., p. 156.

[5] ibid., p. 153.

[6] ibid., p. 155.

Summerlin[1]. Most of these instruments have evolved out of a context of teaching the various science disciplines as year-long courses or as units of a general science course, and therefore are of little use in evaluation of integrated science programmes, except perhaps as sources for items.

In an attempt to reduce the influence of general reading ability upon performance on a science content test, Finkelstein developed the Pictorial-Aural Inventory of Science Knowledge[2] to assess the achievement of American Fifty-graders in science knowledge. It consists of sixty multiple-choice items using pictures instead of words for both item stems and distractors.

The newer curricula have increasingly emphasized the learning of science at the higher cognitive levels: application and comprehension, for example, rather than just at the knowledge level. Helgeson, in studying concept development of American elementary school children, developed an instrument[3] to assess achievement in understanding of certain concepts of force at each of the three levels of knowledge, comprehension and application. Jeffrey, in his case study of the Scottish Integrated Science Syllabus Evaluation (see Chapter 12) discusses the use of instruments which assess achievement at four levels; recall, comprehension, application and highest ability, which were modified from the Bloom categories.

(c) The ability to use scientific processes and to engage in problem-solving behaviour

With the advent of process-centred curricula over the past decade, there has been increased activity in attempts to assess student ability to apply scientific process skills and problem-solving behaviour. This has especially been true at the elementary and junior high or middle school level. As a result, a variety of instruments have been developed to assess such abilities. Fyffe's instrument[4] was designed to assess the process-skill acquisition of elementary school children for the processes of formulating hypotheses and defining operationally, as defined in the AAAS Science - A Process Approach (SAPA) materials. Test of Science Processes[5] was used to assess the abilities of junior high school students to use the science processes of observing, comparing, classifying, quantifying, measuring, experimenting, inferring and predicting. Attempting to minimize the effects of verbal ability on test achievement, Beard developed an instrument which used slides and correlated audio-tapes rather than the written multiple-choice type item in her Basic Science Processes Test[6]. It too was designed to assess achievement in science processes as defined by the AAAS.

Disinger developed a single instrument to assess both concept and process development in Ohio junior high school students. His Concept-Process Test[7] was constructed using as a model the seven basic conceptual schemes and five major items in the processes of science identified in the 1964 Curriculum Committee Report of the National (USA) Science Teachers Association.

(d) Scientific attitudes

A major objective of general education in science has been the development of scientific attitudes; the way in which a student looks at information or situations. Nay and Kozlow

[1] Wall and Summerlin, op. cit.

[2] Mayer, Unpublished Evaluation ... op. cit., p. 70.

[3] ibid., p. 71.

[4] ibid., p. 90.

[5] ibid., p. 88.

[6] Mayer, Unpublished Evaluation ... op. cit., p. 89.

[7] ibid., p. 85.

developed the Test on Scientific Attitudes[1] to assess such scientific attitudes as critical-mindedness, suspended judgement, respect for evidence, honesty, objectivity, willingness to change opinions, open-mindedness and questioning attitude. The instrument was designed for use with senior high school students in Canadian schools.

(e) Cognitive preference

Preferred learning styles of students have received some attention by curriculum evaluators. Atwood[2] developed an instrument to assess the degree to which chemistry courses affected a student's cognitive preference and mode of attending to information, and whether these effects were consistent with the objectives of the course. Tamir, et al, have developed the Agriculture Cognitive Preference Test and the Biology Cognitive Preference Test for use in evaluating the implementation of science curricula in Israeli schools[3]. Despite their other relative merits, each of these assesses cognitive preference relative to instruction in a specific discipline. Atwood went on to develop the Cognitive Preference Examination - II[4] (CPE-II) which utilizes content from general science and social science. Three preferences are assessed: memory or retention of factual information, application of information or what the information is "good for" and questioning or the challenging or criticism of the information presented. The instrument was designed for use with students of the senior high school level. The CPE-II was revised, updated and revalidated for use with college science majors by Wright. Any evaluator interested in using cognitive preference as a criterion should first consult a comprehensive critical review by Brown[5] in which serious questions are raised as to the nature of cognitive preference and the validity of instruments designed to assess it.

(f) Attitudes relating to science

The increased concern among science educators for the role of science curricula in improving students' attitudes toward the science establishment and its workers is reflected in the great number of instruments that have been developed over the past ten to fifteen years.

Champlin recognized a problem encountered by others when developing instruments to assess attitudes. He attempted to construct a scale which would allow the investigator to separate out the influences of knowledge about science and the scientific enterprise from an individual's attitude toward them. His instrument has two parts: "part one" assesses the respondent's beliefs about science, and "part two" measures the respondent's evaluation of those beliefs. Attitude toward science, then, is determined through a complex scoring procedure relating responses on "part one" to responses to items on "part two". The instrument, Beliefs About

[1] Marshall A. Nay and M. James Kozlow, "Test on Scientific Attitudes". Department of Secondary Education, University of Alberta, Edmonton, Alberta, 1973.

[2] Mayer, Unpublished Evaluation ... op. cit., p. 100.

[3] Pinchas Tamir and Moshe Simanovsky, "Agriculture Cognitive Preference Test", De Shalit Science Teaching Center, The Hebrew University, Jerusalem, 1975; and Pinchas Tamir and Rostam Mukades, "Biology Cognitive Preference Test", De Shalit Science Teaching Center, The Hebrew University, Jerusalem, 1974.

[4] Mayer, Unpublished Evaluation ... op. cit., p. 101; and Robert Wright, "Cognitive Preference Styles of College Students Majoring in Science, Mathematics and Engineering", Unpublished doctoral dissertation, The Ohio State University, Columbus, 1975.

[5] Sally A. Brown, "Cognitive Preferences in Science: Their Nature and Analysis", Studies in Science Education, Vol. 2, p. 43-65, 1975.

and Attitudes Toward Science and Scientists Scale[1], consists of thirty-two multiple-choice items in each of the two parts. The scale was used with ninth through twelfth graders from a variety of American socio-economic settings.

The Inventory of Societal Issues[2] was developed to assess the attitudes of high school seniors toward socially significant science-related issues. Using factor analysis, seven interpretable factors were found relating to environmental issues and society and the individual's role in these issues.

A Test of Interests[3] assesses the factors of science orientation, attitude to science as a process and attitude to science as a school subject. Additional factors include interest in science as a leisure process, as a method of inquiring and as a body of knowledge; and non-scientific interest in literature, fine arts and methods of solving problems by consulting books, teachers and specialists. The instrument was developed for use with Australian students in the 10 to 15 year age range.

Brown developed an instrument to assess the achievement of five affective objectives of the Scottish Integrated Science Curriculum. The instrument has five scales (corresponding to the five curriculum objectives), designed to measure: awareness of the inter-relationships between the different disciplines of science; awareness of the relationship of science to other areas of the school curriculum; awareness of the contributions of science to the social and economic life of the community; interest and enjoyment in science; and objectivity in observing and in assessing observations. The approach used in the development of this instrument[4] can be used as a model for instrument development in evaluation studies. Its development and the results obtained with its use are described in Brown's paper (see Chapter 13).

(g) Instruments assessing a variety of student behaviours

Many of the instruments reviewed for this paper were designed to assess several types of student behaviour. An example is one developed by the Schools Council Integrated Science Project called The Special Project Examination in Integrated Science - Ordinary Level (2). It attempts to measure student mental skills, ranging from ability to recall information, to the performance of complex mental tasks including problem-solving. It also assesses attitudes such as willingness, scepticism and concern. The test scheme places emphasis upon the achievement of aims and objectives rather than on the acquisition of knowledge or particular content. It therefore provides a working model of an evaluation scheme for student behaviours especially adapted to the objectives of integrated science teaching. The 1973 version of the examination consists of five sections, two of which include multiple-choice questions, with many based upon graphs and other representations of data. Each of the other sections contains several essay questions. No item analysis information was available.

Instruments that assess climate or process variables

Evaluation efforts in science need to be concerned not only with the results of instruction; data must also be collected which will permit an analysis of the nature of the implementation effort and the degree to which this effort is consistent with the objectives of the curriculum. All too often, little information is collected on such "process" variables. When this happens, failure to achieve objectives of the curriculum is not easily interpretable. Is this failure the result of the

[1] Mayer, Evaluation Instruments ... op. cit., p. 115.

[2] ibid., p. 147.

[3] G.R. Meyer, "A Test of Interests", Jaceranda Press, Milton, Queensland, 1969.

[4] Sally A. Brown, "Affective Objectives in an Integrated Science Curriculum", Unpublished doctoral dissertation, University of Stirling, 1975.

curriculum? Is it the result of the failure of teachers to understand the content? Did teachers have attitudes favourable to the curriculum? Did they use teaching practices consistent with the philosophy and objectives of the programme? A large number of instruments designed to assess such process variables have been developed. Some of the instruments discussed under the previous category can also be used for this purpose.

(a) Classroom activities

Kochendorfer[1] developed an instrument to assess the degree to which biology teachers were using classroom teaching methods which were consistent with those recommended for use in the Biological Sciences Curriculum Study (BSCS) programme. Sagness extensively revised Kochendorfer's instrument to make it applicable for any science class. His Science Classroom Activity Checklist[2] is designed to assess the degree to which a teacher is using methods and techniques that are consistent with the consensus of science educators in terms of effective science teaching methods, methods which facilitate student understanding of the nature of science and the scientific enterprise. Two versions of the checklist were prepared. One was designed to assess the nature of classroom activities which teachers felt should be used for secondary school science instruction. The other determined the nature of activities teachers actually use as perceived by their students. The checklist consists of sixty "yes-no response" statements of activities. Eight subscales were identified. However, a factor analysis performed on a slightly-modified Spanish language version of the checklist completed by secondary science students in Costa Rica, failed to confirm the subtests as factors. In this study, Berty[3] identified four factors: the integration of the use of the science laboratory with the science course, the use of the textbook and reference materials, teacher-centred teaching and the use of problem-solving procedures.

Shay, concerned about student-centred instruction in science, developed the Science Classroom Inventor[4] to be used in assessing individual preference for the student-centred, non-direct science classroom. He developed a criterion model of such a classroom. A pool of items was generated such that each element of the model was represented by at least two items. Four forms were developed of sixty items each: Student Perception, Student Choice, Teacher Choice and Principal Choice.

(b) Interest in the science course

An important index of total class environment during implementation of a science curriculum is that of student interest in the course of study. The general form developed by Silance and Remmers[5] has been widely used in a science-adapted form. Hedley[6] developed an inventory

[1] Mayer, Unpublished Evaluation ... op. cit., p. 168.

[2] ibid., p. 170.

[3] Roland B. Berty, "A Study of the Relationships Between Classroom Activities, Student-Teacher Relationships and the Characteristics of In-Service Secondary School Science Teachers of Costa Rica". Unpublished doctoral dissertation, The Ohio State University, Columbus, 1975.

[4] Edwin L. Shay, "A Study of Relationships Among Selected Teacher Variables and Expressed Preferences for Student-Centered, Non-Direct Science Education", Unpublished doctoral dissertation, The Ohio State University, Columbus, 1974.

[5] E.B. Silance and H.H. Remmers, "An Experimental Generalized Master Scale: A scale to measure attitude toward any school subject". Purdue University Studies in Higher Education, Vol. 16, p. 84-88, 1934.

[6] Mayer, Unpublished Evaluation ... op. cit., p. 131.

consisting of seventy-two statements with a Likert-type response format to assess student acceptance of course material, course content, laboratory work, interest in the course, involvement, and satisfaction of perceived needs. Reliability and validity estimates were not reported in its use with tenth-grade Manitoba students taking the "General Course" science programme. More recently, the semantic differential format has been widely used, such as in Henson's study[1] of science achievement of students using the Earth Science Curriculum Project materials in Oklahoma City schools. His scale attempted to measure attitudes of students toward school, science, learning earth science, reading earth science, earth science experiments, earth science teacher and earth science classmates. His semantic differential used those seven concepts, each with nine adjective pairs as the response scale.

(c) Teacher variables

Another important dimension of the nature of the implementation effort consists of the background knowledge of the teacher and teacher attitudes and opinions concerning appropriate instructional objectives. Unless these factors are consistent with the objectives of the curriculum being implemented, it is doubtful that an optimal classroom climate can be achieved. Mayer, et al[2], developed an instrument to assess the achievement of teachers enroled in the summer phase of a workshop designed to prepare them to teach the Earth Science Curriculum Project (ESCP) materials. Their *Earth Science Test* was prepared using as a reference the behavioural objectives as stated in the "Teachers Manual" for *Investigating the Earth*, the ESCP student textbook.

Merkle designed[3] an instrument to assist in determining the effectiveness of a leadership workshop on elementary science. It was designed to test knowledge of the programme characteristics of AAAS *Science - A Process Approach* and of Science Curriculum Improvement Study.

Wonkka[4], in evaluating in-service institutes for science teachers, used a semantic differential of twenty-five concepts and sixteen bipolar adjective pairs to determine teacher attitudes toward the "new" curricula. James[5], in working with student teachers in science, used a semantic differential of twelve concepts and fifteen bipolar adjective scales to assess attitudes toward teaching science. Validity was established through use of a factor analysis of results.

Schwirian developed the *Science Support Scale*[6] to be used with undergraduates in teacher education and elementary school teachers to assess their attitudes toward science. Items were selected and developed in accordance with Bernard Barber's contentions regarding the nature of science attitudes necessary to the growth and the development of science in a society.

Berty, in studying science education in the secondary schools of Costa Rica, developed the *Rating Scale of Objectives of Science Education*[7] to determine what objectives teachers were emphasizing. He reviewed stated objectives from the ministries of education of five Latin

[1] ibid., p. 134.

[2] Victor J. Mayer, John F. Disinger, and Arthur L. White, "Evaluation of an In-Service Program for Earth Science Teachers", *Science Education*, Vol. 59, p. 145-153, 1975.

[3] Mayer, *Unpublished Evaluation ...* op. cit., p. 175.

[4] ibid., p. 178

[5] ibid., p. 172.

[6] Mayer, *Unpublished Evaluation ...* op. cit., p. 121.

[7] Berty, op. cit.

American countries, identifying forty-four objectives. Respondents rated each on a four-interval scale from "Very much emphasis" to "Very little emphasis".

CONCLUSIONS AND RECOMMENDATIONS

The type of data-collecting procedures discussed in detail in this paper is only one of several formal and informal types that have been used in curriculum evaluation studies. The reader is warned, therefore, not to presume that the formal data-collecting procedures discussed herein are the best for use in any situation, or that instruments developed to assess a student behaviour resulting from one situation or curriculum can be used intact with another situation or with another curriculum.

In formative evaluation, most workers have found interviews, classroom observations, open response forms and other informal methods to be more useful in providing the detailed information necessary for effective curriculum revision. However, information on student characteristics which should be used in making decisions in formative evaluation can perhaps best be obtained through the extensive use of the types of instruments reviewed in that category in this paper. The types of procedures described here are particularly useful in summative stages of evaluation where there is some attempt to compare the results of one curriculum with those of another. For this purpose, a higher degree of objectivity of data and data collected from a much larger sample or population than is possible with informal means, are desirable. A widespread testing of the curriculum in a wide variety of contexts and the provision of evaluative data from such testing can provide essential information for decision-makers in selecting curricula or curriculum elements for implementation. This type of data can probably best be supplied through the use of objective, respondent-completed instruments.

A final warning: the reader should not take the description of specific instruments for assessing student performance and climate or process variables as an endorsement of those instruments for use intact elsewhere. For most purposes, an instrument must be designed to measure the specifically-intended outcomes of a curriculum and the procedures and philosophy of that curriculum. These instruments are described as models that could be used by an evaluator to develop similar instruments for his or her own purposes. It should be emphasized however, that instrument development is not a simple procedure. To develop sufficiently reliable, valid, and objective instruments requires heavy investment of time and creative energy. Unless sufficient effort and expertise are applied to instrument development, the information yielded by their use may be meaningless, or worse, misleading.

At least in one area of instrument development, it might be useful to redirect our efforts. We are at a point where there is a wide range of attitude instruments available, most purporting to assess similar aspects of student attitude toward science and scientists. Effort should now be directed toward the revision and/or refinement of these instruments along the following guidelines:

The specification of a clear theoretical construct to underlie the instrument. This construct should then guide the selection and development of items.

Avoidance of confusion between different theoretical constructs. If more than one is to be included in a single instrument, each should be identified and scored as a separate factor.

Provision in the design for filtering out influences of respondent knowledge about the scientific enterprise from attitudes towards it.

The elimination of defective items, such as those that combine two or more different understandings or perceptions.

The preliminary trial of the instrument on a population with characteristics approximating those of the population with which the instrument is to be used.

The refinement of the instrument for each factor or subscale such that a reasonable internal consistency is obtained. (This relates closely to the first two items.)

Determination of the stability of the instrument through a test-retest technique.

The use of factor analysis to empirically validate factors.

A few attitude instruments developed and refined in this manner will prove much more useful than the continued generation of unrelated instruments in isolated studies.

The assessment of gains in science concepts and processes and in knowledge about the scientific enterprise will, and probably should, continue to be the prime focus of evaluation efforts. Since the content of integrated science courses will be highly dependent upon the particular curriculum, in that it must reflect local conditions of culture and environment, it will be fruitless to attempt to develop instruments that can be used extensively. Instead, instruments have to be assembled for the specific curriculum in question. The development, testing and refinement of items, however, is a difficult and expensive process. Also, even though instruments must be specific to the local curriculum, individual items would have wide evaluation utility for curricula throughout the world. Therefore it is recommended that item-banks be developed and made available through an international agency such as Unesco. Past evaluation efforts have produced an extensive variety of items. These items should be collected and classified according to some conceptual framework. One that might be adapted for this use is that developed by the Federation for Unified Science Education[1]. It consists of four categories: the unifying concepts of science, persistent problems, processes of science and values of science. Extensive effort has gone into the development of item-banks in the areas of biology and physical science under the auspices of the Australian Council for Educational Research and the Curriculum Centre of the Education Department of Tasmania. Such an item-bank could eventually provide standardized items on a world-wide basis on the major common aspects of science. It would not preclude the local development of additional items relating to instructional objectives that may be peculiar to the local situation because of cultural and environmental factors. The availability of such an item-bank would facilitate what has long been the ideal of achievement testing: use of domain-referenced testing in which items are randomly selected from a large bank (domain) that has been developed with reference to all the objectives of the curriculum being evaluated. In the past, this has been impractical because of the amount of effort and energy necessary to create a domain of a thousand or more items required for an adequate representation of a curriculum.

Most of the available instruments have been developed to assess aspects of student performance. As we have become more sophisticated in our efforts at curriculum development, we have come to realize that curricula need to be highly flexible to allow for the variety of student characteristics which are related to effective learning. To develop such curricula, however, we need to be able to identify important characteristics and to be able to measure them in order to evaluate the relative effectiveness of curriculum elements. Efforts such as that by Tisher and Dale, in developing instrumentation for identifying the Piagetian stage of development relative to physical science concepts, are encouraging. This needs to be extended into other science concepts. Effort must be directed toward identification of out-of-school experiences which relate to science achievement and to the development of instruments accordingly. Related to this is the effect of culture on science learning. Griffiths and Smart[2] comment on this problem relative to the development of an interdisciplinary science course for Papua New Guinean students. Their experience demonstrated that there are some scientific concepts and much practical experience implicitly assumed in most western science courses which were not necessarily present in the background of their students. On the other hand, certain scientific modes of thought, particularly classification, have been found to be highly developed in students in certain non-western cultures. Such concerns become especially important in developing science curricula within highly

[1] Center for Unified Science Education: Prism II, Vol. 3, No. 1.

[2] K.G. Griffiths and R. St. C. Smart, "An Experiment in Interdisciplinary Science Teaching-Preliminary Year Science, University of Papua New Guinea". Science Education, Vol. 59, p. 27-38, 1975.

pluralistic societies, and in such situations at least, it would seem important to have instruments which would identify student differences arising from cultural factors. There are no doubt other areas of student background characteristics that must be considered as well, such as critical-thinking behaviour in science. This is where the future emphasis upon instrument development should occur.

It is encouraging to see the development and use of instruments relating to classroom climate variables as they occur during the implementation effort. The work of Kochendorfer and Sagness in the United States and Best in Australia, using various forms of a Classroom Activity Checklist, and Walberg and Anderson in the United States with the Learning Environment Inventory[1], have been pioneering efforts in science education. Their efforts, and those of others using their instruments, have demonstrated that the student is a highly reliable observer of classroom practices. Future instrument development should capitalize on this fact and extend their work to a wide variety of curriculum-dependent, learning-environment characteristics. The collection of this type of information is extremely important in judging why a particular implementation effort was either successful or unsuccessful and in describing for the decision-maker the climate in which a given curriculum can achieve success.

The case study by Sim of the evaluation of integrated science teaching in Malaysia includes the description of a broad-based data collecting procedure containing many of the elements recommended in these conclusions. The reader is referred to Chapter 15 for a description of the types of variables that are being measured and the data gathering techniques that are being used in that study.

BIBLIOGRAPHY

Balzer, A.L., Evans, T.P., and Blosser, P.E.: A Review of Research on Teacher Behaviour Relating to Science Education, Association for the Education of Teachers of Science and the ERIC Science, Mathematics and Environmental Education Clearinghouse, Columbus, Ohio, 1973.

Contains a review of classroom observation instruments that have been developed for use in science education.

Gardner, P.L.: "Attitudes to Science: A Review". Studies in Science Education, Vol. 2, p. 1-41, 1975.

A review of some instruments developed for assessment of attitudes to science.

Wall, Janet and Summerlin, Lee: Standardized Science Tests: A Descriptive Listing. National Science Teachers Association, Washington, D.C., 1973.

Descriptions of 51 achievement tests developed by commercial publishers.

[1] Gary J. Anderson and Herbert J. Walberg, "Learning Environments", Chapter 6 in Herbert J. Walberg (ed), Evaluating Educational Performance, McCutchan Publishing Corporation, Berkeley, California, 1974, p. 81-98.

7 Assessing student progress in integrated science using teacher-prepared methods

Gregor Ramsey
Torrens College of Advanced Education, South Australia

SUMMARY

In this paper on assessing student progress in science using teacher-prepared methods, the emphasis has been placed in two main areas. The first is the conceptualization of assessment in terms intended to be useful to the teacher. The second is the clear identification of assessment of students as much more than pencil-and-paper tests and the outlining of a few areas which teachers may explore.

The nature of assessment

Despite the noble aims of integrated science courses and statements made about how they should be taught, the manner of assessment and the content of the assessment instruments reflect what the actual curriculum is supposed to be about most accurately for students. Their real science curriculum is defined by what is assessed, because they know that, based on this assessment, they either pass or fail, and their future is in the balance. For this reason, one can see the importance, in the framework of integrated science curriculum, of the way the students in the course are assessed by the teacher.

At this point, we should define our terms. It is easy to use the word assessment and assume everyone knows what it means, or to slide from the word 'assessment' to the word 'evaluation', assuming the two to be synonymous. We may also describe a student's 'performance' on a particular test and then draw 'inferences' from such performance about the learning abilities of a student. We may establish 'criteria' for passing or failing a student, or for granting a credit rating. Eventually, we may have to 'report' to the student, the headmaster, the parents or an employer the results of our assessment.

Some simple definitions which have been found useful by groups of science teachers in South Australia for assessment discussion may prove helpful:

performance	-	is any activity which may be observed or reacted to;
inference	-	is any statement about the unknown based on the known;
judgement	-	is any statement of approval or disapproval which generally closes further thinking;
assessment	-	is a description of a performance and may be descriptive of fact, inferential or judgemental in emphasis;
criteria	-	are stated aspects of performance on which an assessment is based;
reporting	-	is the process of communicating an assessment.

The term 'evaluation' may be considered in restricted definition to be a statement about the relationship between performance and criteria, and so all evaluations are assessments.

An important role of the science teacher is to attempt to find out what students are thinking and how they are learning from a wide range of external performance cues shown by them.

An example of such a cue may be a simple exchange in the middle of a lesson, such as:

performance:	A student interrupts the teacher in mid-sentence with a loud noise.
inference (by teacher):	The student is testing me.
judgement (by teacher):	The student is rude and should be punished.

Such interchanges have profound effects on classroom learning. A second example:

performance:	A student interrupts and asks a teacher why a person can drink vinegar if it is an acid.
inference (by teacher):	The student has not used information he should have on strong and weak acids and dilution.

From the inference the teacher may either make a judgement and hence closing thinking or decide a course of action. For example:

judgement (by teacher):	The student has not been listening in class (producing closure).
course of action:	The student needs some help in understanding acids (options remain open).

These sequences of

student performance
teacher inference
teacher judgement

or student performance
teacher inference
student and teacher's course of future action

are common in a classroom. They may arise from verbal interchange or from the student's performance on a test.

Assessments are usually of at least three different kinds. There are (a) descriptive, factual summaries; for example, on a test a student could match only three scientific terms correctly from a group of ten; (b) inferences drawn from the descriptive summary; for example, having achieved only three correct matching of terms, the student has not learned the terms. This leads to the action of advising the student to spend more time learning the terms; and (c) judgements made on the basis of the inferences; for example, the student was too lazy to learn for the test. Although the assessment of performance is often a complex intuitive business, it may be helpful to consider it in stages:

observation of the performance
↓
factual summary
↓
inference
↓ ↓
judgement | course of action

When being assessed by a teacher, students sometimes do not appreciate the distinction between these stages or types of assessment. In fact, the inferences they may draw from the facts of the performance, or the judgement they may make, may be quite different from those made by the teacher.

From the teacher's perspective, the set of events may be described as follows:

observation of performance (agreement by student and teacher):	Student achieved three correct out of ten.
teacher inference:	The student needs to spend more time learning.
teacher judgement:	The student is lazy.

From the student's perspective, the set of events may be described as follows:

student inference:	The teacher did not explain the terms well enough for me to understand.
student judgement:	He is a bad teacher and I can never learn with him.

If the student's inference and judgement happen to be the same as the teacher's, then an agreed course of future action by student and teacher is likely to have some success. If, however, there are widely divergent inferences, drawn either consciously or unconsciously by the student as in the example, then likely future courses of action are going to be much more difficult to achieve. Any future course of action will require the teacher to use positional power as an authority figure for there to be much hope of the course of action producing an improvement in the student's level of learning. When student and teacher agree either overtly or covertly in their inferences and assessments, co-operation in future between student and teacher is possible. When they do not, then co-operation becomes much more difficult and the teacher is likely to need to put the student under duress to force learning.

Criteria for assessment

Much of what has been stated so far has been based on intuitive understandings of the teacher or the students about their achievement as a result of the learning process. If we openly state what we expect a performance to be before we observe it, then we are establishing criteria.

Criteria for assessing performance on actual science learning may be specified. Such criteria may be provided by a checklist that may be used by the student, by another student, or by the teacher to assess (say) whether the student can use a balance properly. Another example of such criteria may be for the teacher to establish that a report of a field experience shall have a certain number of words, will identify the natural habitat of five birds seen in the forest and contain diagrams of five different leaf forms. When criteria are stated, then both students and teacher know what is expected. If these criteria are not achieved by all students, then judgement may be made on any factors which may be considered to affect the achieving of the criteria.

The relationship between actual performance and the established criteria may be of at least three kinds:

comparative	-	for example, the teacher says 'this student's report is well below the average for the class', or 'your behaviour in practical classes is much worse than the other classes I teach';
diagnostic	-	for example, the teacher says 'this report does not contain a summary or conclusion about the kinds of beaks that the birds in that part of the forest have', or 'the bad behaviour of the class seems to stem from you three working together. I think we will separate you';

predictive - for example, the teacher says 'if you do not learn to put in summaries and conclusions in your reports then it is unlikely, if you ever become a scientist, that your work will be useful to others', or 'if this class continues with this bad behaviour, then many of you will fail at the end of the year'.

The function of assessment

Generally in classrooms, it is assumed that the comparative, diagnostic and predictive assessments are made by the teacher. Usually this is so. Yet if the students are prevented from being involved in these processes in the classroom, they will make their own assessment themselves, either covertly in class or with their peers outside class time. Since one of the generally-stated aims for integrated science is to foster a student's decision-making ability, it seems important that students themselves as well as the teacher, be actively involved in the evaluation process.

We can assume that most students have hopes and aspirations for themselves which lead them to want to know: whether they are capable of and are suited to what they want to do; whether what they are doing is worthwhile, either in the present context of schooling or in the future context of employment; how to improve what they are doing so that they can do it as well as they are able; and how well what they are doing compares with how well others do the same or a similar thing; or, does their performance meet some minimum acceptable criteria for the task or activity.

The first of these assumptions relates to our providing better information for students about the science they are capable of learning, and indirectly to the teacher about teaching. Integrated science currently seems to be what teachers of integrated science want to, or are given to teach. How do we decide what is appropriate for the group of students in the class, given the resources available? The criteria for deciding this information ought, if possible, be agreed upon by students and teacher, either by a quite formal process (especially in the senior school) or by a much less formal approach (in the lower grades). It ought to be possible to help students understand how the particular topic chosen suits their present interests and level of knowledge, skill and ability.

The second assumption relates to the worth of what we have the students do. If the students value what they are learning, then they are more likely to want to learn it. If the teacher can communicate to the students why a particular topic or set of learning experiences is valued, then the students are more likely to value it too. Evaluation is about values, both of the teacher and of the student. The very process of evaluation has its effect on the development of students' attitudes to science and to learning. And the way the teacher assesses and judges will reveal the teachers' own value system to the students.

The third assumption relates to improvement. When criteria have been established and agreed upon by student and teacher, evaluation lets the students know how far they are from meeting these criteria. Assessment against overt criteria informs students at one of three levels:

1. That they do not know a particular piece of information or cannot do a particular task; the pass-fail kind of information.

2. What they do not know, or which tasks they cannot do; such information identifies problem areas, but does not help the students know how to go about improvement. For example, the teacher states, 'your knowledge of the structure of the human heart was sound, but you did not understand the mechanism by which the heart continues to beat'.

3. What they can do to improve what they do not know; here the emphasis is on diagnosis and remedy with the assessment being part of the instructional process. To continue the second example, specific pages of a book may be indicated to the student or he may be asked to follow the matter up individually with the teacher later.

We are confronted with a hierarchy of assessment at three different levels. The first kind of

information can be set against criteria, some of which may be unknown to the students. Public examinations are an example of this kind. If students do not know the relevant criteria, then they cannot make inferences which will help them take a course of action to improve. The final grade or mark given in a public examination is a judgement and produces closure. The second kind of information provides some information about the criteria and is more open about the teacher's evaluation against criteria; but if this is not to produce closure, then future action is largely up to the student. The third kind of information provides the criteria, how close the students come to meeting the criteria and what the student needs to do to meet the criteria eventually.

The fourth assumption relates to the genuine desire of students to know how they compare in their learning with others. Information of this kind is provided by ranking students on a particular set of tests. From the distribution of marks and the student's rank, we can assign grades. The assessment criteria are comparative, relating to some 'average' student in the class. Such comparisons can be important to the students, their parents and the teacher, but can reinforce failure for those students who are always below average. The challenge for the teacher is to establish the appropriate balance between comparative criteria for all students and absolute criteria for each individual student.

Our tests in the classroom should help us as teachers to obtain information about our students, and help the students to obtain information about themselves so that the teacher's evaluation of the results and that of the students lead to judgements or courses of action from both which are as congruent as possible. Tests should be looked on as aids the teacher has for helping students make decisions about themselves, as much as providing opportunities for teachers to make judgements about students.

The nature of a test

A test provides a method of sampling the performance of students. It may sample knowledge, skills, attitudes; but it can never be more than a sample. The more representative the test is of the learnings expected, then the more accurately will the test sample reflect the total learning of each student. Of course, the teacher's perception of how accurately the test samples expected learning may be quite different from the student's perception. The more open the criteria for assessment of performance to the child, the more likely are the student's and the teacher's perceptions to coincide. And to the student, it is what is sampled that provides the evidence for what the teacher sees as important in the curriculum. It is sometimes a salutary experience to ask a class to set a test on a section of work, perhaps for homework. Despite their undoubted lack of sophistication in test setting, what is obtained does show how very differently students in the class perceive what may be expected.

The test as a sample selected by the teacher from a particular science topic should take into account: the aims of the course; the special interests of the group of students; the ability of the group of students; the time available for the test; those things which may be necessary pre-requisites for the next section of learning; the resources available; and the method of expression of the ideas.

The word 'test' may be used to describe any device or method used to sample students' behaviour and may or may not be of the pen-and-paper kind directed at the individual child.

Main uses of a test

A test may be used as a pre-test which measures pre-requisite knowledge and skills that a student has before beginning a learning sequence. Such tests are useful in preparative evaluation, where the teacher makes an assessment of students' abilities, interests and knowledge prior to the beginning of instruction. The teacher establishes what is expected as entry criteria and then tests to see how closely the particular group of students meets these criteria. Should the students not meet the criteria, preliminary instruction may be given to those students falling short of the anticipated behaviour. For example, a particular instructional sequence may require the

student to use a microscope. A simple pre-test would show those who meet criteria for success in using it and those who do not. Those who cannot may then be given special instruction in the use of a microscope.

Tests are used to help the teacher assess how effectively student learning is occurring during the instructional sequence. This may be done by checking simple homework questions, asking students what difficulties they are having with the work or watching to see whether the students are undertaking purposeful activity in the laboratory. This is very much a matter of experience and leans heavily on informal monitoring by the teacher of what is going on. Often the difference between a 'good' teacher and a 'not so good' teacher is in this constant process of intuitive monitoring, and it is an area in which much more help ought to be given to teachers.

If the teacher senses that something is specifically wrong in the instructional process, then diagnostic tests may be used in an attempt to focus on the difficulty. Teachers can quickly develop simple diagnostic tests. They should be brief, direct, and in such a form that, wherever possible, the students in discussion with the teacher can diagnose their own problems. Diagnostic tests are used for checking simple skills; for example, using a balance, pipette, scales of various kinds, slide rules or log books: for checking basic knowledge such as symbols, formulae and equations, or terminology, say the parts of the body, or simple demonstrations of the understanding of basic laws. Often, in-depth interviews with students about their work can provide diagnostic information which is superior to that obtained in other ways.

Finally, tests are used to assess what students have learned in the total instructional programme. These tests are generally used to compare student levels of attainment with each other, to provide rank scores and who is to be 'top of the class' in the particular test. These tests may be provided by external examiners, or as separate test batteries which are part of the integrated science materials being used. Knowledge can be tested quickly if so-called 'objective', multiple-choice, pen-and-paper tests are used. Reports of work done, and discursive answers to specific questions are also used. Much more rarely oral reports either to the class or to the teacher, or in-depth discussions are used to make evaluations. Criteria for assessment of such non-written methods are much more difficult to state, but they can give a better total picture of learning than pencil-and-paper tests, particularly when outcomes other than knowledge are expected.

Criterion- and norm-referenced tests

There are two basic kinds of tests which teachers can prepare. The first are criterion-referenced tests where student performance is measured against some pre-determined standard of reference. Such standards for an individual student do not directly relate to the performance of other students, but, of course, must be based on a realistic expectation of what may be achieved. Such tests, sometimes called mastery tests, are used generally as pre-tests or formative tests or diagnostic tests. They may also be used as summative tests, but it is usually at the end of a learning sequence that comparative data are required. With such tests, the students would know that they would have to be able to identify, say, eight species of plant out of ten with a key provided, before it would be assumed that sufficient 'mastery' of a topic on identifying plants had been achieved. In such circumstances, some students would achieve mastery much more quickly than others, and hence their use is often based on a more individualized approach to learning.

Teachers should be encouraged to develop this kind of test more often. They state clearly for the student the performance criteria expected and orient the student toward success, rather than failure.

The second general type of tests are norm-referenced tests. These relate the performance of a student to that of other students in a class, or other clearly defined group. What is acceptable depends on the relative performances of one group of students as compared with another. They tend to use competition as a major motivating factor. They are important for making performance comparisons, for example, to show whether a particular criterion-referenced test is too severe.

Different contexts for assessment

Although two very specific types of tests have been given, the concept of assessment in an integrated science classroom must have a much wider perspective than simply obtaining information from pencil-and-paper tests. Teachers are encouraged to explore the different contexts in which assessment may be made and the different media which may be used to vary the presentation of assessment material. Contexts which may be explored are groupings, places, time and structure.

For grouping, we may assess individual students, small groups acting as teams, say in discussion sessions or practical sessions in the laboratory, and class-size groups where we may wish to compare one class with another or simply determine how the class is progressing. One technique which gives a considerable amount of information about a class group within a reasonable test duration is to divide the class randomly, say into four groups. Give each of the random groups a different test. From this arrangement, it is possible to obtain four times the information about the class in the same amount of time. It does not, of course, give the same specific information about each individual student in the class.

Assessments may occur in many places, not just in the classroom. There is the laboratory for tests of practical skill and problem-solving. In the field, on field excursions, students may be expected to identify certain animals or plants; to sketch particular geological formations or to look for examples of man's impact on a particular environment. Evaluative discussions ought to form an important part of any field exercise.

The home, too, can provide a place for assessment. Parents, if possible, may be involved in what is done at home. Students may be given checklists and diagnostic tests which may be taken at home. In many settings, however, very little can be achieved in the home environment.

The library should be the place where information search skills are tested. For example, whether a student can use a library card index, the index of a book, a data book of information about the elements can all be tested in the library. Some tests of this kind have been simulated in pencil-and-paper form, for example, by Fraser[1], but it is by far the best for the students to be taken to the actual library.

The school grounds and neighbourhood also provide an interesting and different context for assessment. Students may be given checklists which would assess whether they can apply what they have learned in the classroom to a real situation.

The time taken for any assessment needs to be assessed carefully. In general, students should enjoy the experience as much as any other set of learning experiences put before them. This is much less likely in the competitive atmosphere of normative testing than in explicit criterion-referenced assessment. Generally, too, teachers over-estimate what can be achieved by a student in a given time, particularly when the time required to read and comprehend a particular question is taken into account. Summative assessments which give information at the end of a particular instructional sequence may be weekly to annually, or at whatever seem convenient intervals. Other forms of assessment should be initiated as the need arises.

The use of time by students is an important aspect to be assessed. How much time is lost at the beginning of a lesson? How long do particular activities take? When should a discussion be shut off? These are all aspects of student use of time which should be monitored wherever possible against criteria.

Teachers should give some thought to the structure and form of their test. Highly structured tests, for example, the so-called objective type, allow specific identifiable performances to be observed and the result of the test put in number form. For this purpose, we can use multiple choice or matching pair or true/false types which give little flexibility as to which answer is correct. What the teacher expects as the correct answer is among the possible responses and there is little room for the 'happenstance' different response of some students. More open responses from students may be encouraged by having the assessment less tightly structured; for example, students in a group are invited to discuss a question like 'What difference does the

[1] B.J. Fraser, Tests of enquiry skills, Books 1-3, Monash University, 1974.

use of fertilizer make to agricultural production?'. While certain responses may be predictable, others may not. It is in such more open situations that student attitude development may be assessed. In the question given, if the integrated science course had as one of its aims to 'make students aware of the impact of man on the natural environment, we would hope to see students raising issues related to human management of the environment.

Much more weight should be given in classrooms and laboratories to observations of unplanned student performance, especially for monitoring the development of attitudes. Teachers should keep a 'daily log' of important behaviours of students which may relate to teaching objectives. Such anecdotal records are extremely valuable in helping assess the instructional process and in providing formative evaluation data. Specific difficulties students have with their learning might be listed; records of tasks which required too high a level of skill will be invaluable for the teacher in deciding what students might attain.

The medium of assessment

When considering the medium or vehicle of assessment we should recall that student responses may be of three kinds: written or spoken, or an action not requiring words. Thus in deciding whether a student can prepare a salt from hydrochloric acid and sodium hydroxide, we may ask for written evidence or an oral report to the class, or we may ask, given the two solutions, to prepare a dry sample of the salt. If we look at criteria which may pertain in the three cases, we can see that the three different types of performances allow us to assess different outcomes using the same piece of content. These may be summarized as in Table 9.

Table 9. Assessment of different outcomes using different assessment methods.

Written	Spoken	Action
Individual activity	Individual or group activity	Individual or group activity
States how	Says how	Shows how
Written expression skill paramount	Verbal expression skill paramount	Manipulation skill paramount
Interaction between individual and paper	Interaction between individual and peers and teacher	Interaction between individual and materials
Feedback re success or failure delayed	Feedback re success or failure immediate	Feedback may be immediate or delayed
Recall from remembered events important - cueing by teacher difficult	Recall from remembered events important - cueing by teacher easy	Recall may be cued by concrete objects and by teacher

Whether one expects a written, spoken, or action performance from the student depends on a number of factors. These include:

- feasibility - sometimes only written or spoken responses are possible because of resources.
- student age and stage of development - in general, more highly verbal responses are suited to highly verbal children. Younger children are better suited to 'action' tests.

outcome being tested - if we want to know whether students really can make salt, action is the only way.

time available - spoken report can take five minutes, a written report fifteen minutes and an actual demonstration a whole lesson.

The important thing is that all three methods should be used in assessment. Some examples of the three types of student performance are given: teachers are encouraged to explore the best medium for the particular task.

Examples of written tasks include:

use of pen or pencil and paper to report or respond;
overhead transparencies prepared by student;
open book or closed book tests which may be -
- multiple choice
- short answer
- true/false
- matching pair;

essay topics which may be prepared or unseen;
written reports of student stating -
- what has been learnt
- what can be done which could not be done before
- what is found difficult
- what is found easy
- what is found interesting
- what is found boring;

practical records;
student laboratory practical records;
student projects;
drawings and sketches;
report of a significant (to the student) investigation to be read by -
- the teacher
- other students
- parents;

written responses to situations shown in or main ideas from -
- television programme
- film or film strip
- paper report.

For some of these tasks, the student needs to be guided by fairly closely stated criteria; this is particularly true of reports of laboratory work or student projects. Other tasks are more open-ended where the teacher uses the sum of the responses of the class to assess how effective teaching has been. For example, if on asking the students what they find difficult the teacher finds more than fifty per cent identify valency and formulae, more class teaching time may be set aside for it. If only one or two do, then individual work can be provided for the one or two and the teacher can go safely to balancing equations.

Examples of spoken tasks include:

oral answers to written questions, for example -
- student explains to class an answer to a question at end of chapter,
- student answers to class a question selected at random from a box of questions;

oral answers to oral questions -
- asked by teacher
- asked by other students;

discussion/interview -
teacher with individual student
teacher with small group of students
student with student(s) and teacher observer;

audio and video tape -
made by student and discussed by students
made by teacher and discussed by students
from outside source;

reports to class -
of an experiment completed
of library search attempted
of a field excursion
of some unusual phenomenon;

questions asked in class for oral response -
by teacher of students
by students of teacher
by students of students
(the first of these is very common, the second common and the third rare);

discussion on -
written projects
practical situations
television or film programmes.

Criteria for assessing spoken performance are less well identified than for written performance responses and for this reason spoken performance is seldom evaluated. It is interesting occasionally to tape record a small group or class session to see which students in the group: show they have grasped the content; can articulate ideas on the topic; ask good questions; contribute to the cohesiveness of the discussion; and create disruptive events.

If such tapes are made at regular intervals, some idea of class development of oral skills may be gauged. Simple checklists in class to note how many responses a day a particular student makes and to what effect, are also helpful.

Some examples of non-verbal action performances of students include:

students undertaking laboratory activities;
students showing other students how to do something;
students manipulating equipment including -
laboratory equipment
video equipment
audio tape recorders
film - movie and still camera and projector;

students sketching -
on field trips
in laboratory;

students attending regularly and punctually;
playing simulation games;
practical tests of laboratory performance;
students showing particular behaviour patterns on field trips, for example, reflecting conservation attitudes.

The distinctions between writing, speaking and doing are arbitrary and some of the best assessment activities for students require some kind of amalgam of all three. Written assignments are generally most easily converted into marks and hence into grades because many tests (for example, the so-called objectives tests) are easily scored; and somehow as teachers,

we are more comfortable giving a score out of ten to a written statement than we are to either its spoken or demonstrated counterpart. For example, we would be comfortable assessing the written description of how to make salt, but much less so to assess either the spoken description or the actual practical making of the salt. This is mainly a matter of practice on the part of the teacher and of being prepared to try to establish relevant criteria for non-written student performances against which we may evaluate our students.

Methods of assessing performance

In most integrated science curricula, we are looking to assess four broad kinds of performance of students. They almost certainly overlap, and it may be useful to explore the four briefly. The first broad area includes the whole range of methods which may be applied to test knowledge or cognitive outcomes. Teachers in general are good at preparing these tests and there is considerable help available. A good source is Bloom, particularly the chapter by Klopfer on Evaluation of Learning in Science[1]. This source is particularly useful for developing objective tests where the test item is related to a specific objective. The final test may then be either norm-referenced or criterion-referenced depending on purpose. A simple reference which refers specifically to criterion-referenced tests is Gronlund[2].

When preparing an instructional unit, the total sequence of "objective → process of achieving objectives → assessment of actual outcome compared with objectives", should be considered together. Consider, for example, a topic on metals. A general aim for such a topic may be 'to develop in students a knowledge of common metals in use in the community, how they are obtained, their properties and uses'. The students should be aware of the importance of metals in the community and have some appreciation of the different applications of metals. Knowledge, skills and attitudes are all part of the general aim, and as teachers, we should expect any progress to be reflected in student performance and hence be assessed. A total programme of instruction to achieve this general aim may be broken down as in Table 10.

1 B.S. Bloom, J.T. Hastings and G.F. Madaus, (eds), Handbook on formative and summative evaluation, New York, McGraw Hill, 1971; L.E. Klopfer, "Evaluation of learning in Science, Chapter 18 of ibid.

2 N.E. Gronlund, Preparing criterion referenced tests for classroom instruction, New York, Macmillan, 1973.

Table 10. Analysis of a programme of instruction[1].

Intention	Process	Assessment	
Specific objective	Activities	Specific testable performances	Other observable performances
Metals are found in the ground as ores which do not necessarily look metallic. The names of simple ores to be known.	Samples of ores are examined with hand lens.	Correctly name eight out of a group of ten metallic ores when handed samples. Be able to use words like ore, mineral, mining and extraction correctly in written sentences.	
Chemical processes to extract metals from ores.	Reduce some oxides of metals on a match head.		Students describe to teacher which oxides would reduce.
Metals have played a significant part in the development of civilization.	Read books on the relation between metal discoveries and community technology.		Group discussion of effect of metals on civilization.
And so on.			

If we prepare a table of this kind, we can see which of our expected outcomes may be tested by pen-and-paper objective tests and which can best be tested using some simple observational techniques.

If we take the following kinds of objectives seriously in our integrated science courses - evidence in support of claims; rational explanations; observations repeated honestly; co-operation; and concern with the consequences of technology - we are not looking less for outcomes than to test ways of observing what the students are doing. They describe the ways it is intended that the students should learn or perform while learning. So they are best assessed by evaluating the processes whereby learning is occurring. Cognitive and skill objectives may be assessed as outcomes at the end of a sequence; attitudinal objectives are better assessed while the students are learning. Some examples[2] of attitudes and student behaviours which may be observed while learning are shown in Table 11, over page.

[1] Based on Australian Science Education Project, Metals From Their Ores, ASEP trial version, Melbourne, 1973.

[2] Based on material in Education Department of South Australia Secondary Science Curriculum Committee, Do-it-yourself Curriculum Guide for Junior Secondary Science (trial edition), Adelaide, 1974.

Table 11. Examples of attitudes and student behaviours.

Attitude	Student behaviour consistent with attitude	Student behaviour inconsistent with attitude
Awareness of purpose of experiments.	Can give outline of procedure for experiment relating to its purpose.	Can only proceed if instructions are given.
Respect for data.	Records own observations, not what he thinks ought to be observed.	Aaks what the result 'should' be.
	Records unexpected results.	Records what 'ought' to be seen.
Rational interpretation of results.	Accepts 'no conclusion' as a possible outcome.	Makes sweeping generalizations on inadequate evidence.
Desire to understand basic principles.	Reluctant to use a formula he does not understand. Asks questions when this happens.	Student finds out 'the formula' and uses it to get answer.
Persistence.	Continues investigation until complete - tends to ignore end of lesson. Repeats parts of experiments 'to be sure'.	Loses interest in experiment before complete. Finds out what happens from others.

The teacher may extend these and also explore which teacher behaviours would encourage the appropriate responses in students.

The development of <u>skills</u> in students may best be assessed by direct observational methods. In integrated science courses, the notion of 'enquiry' is dependent on the students having the skill to undertake appropriate investigations. Often we expect students to develop skills indirectly. Such skills include: following specified procedures; recording data accurately; using specific pieces of apparatus; and finding information from reference sources.

Many skills important in integrated science investigations must be taught directly. Such skills include: using a cathode ray oscilloscope; using a classification key; and using correct levels of accuracy of measurement.

Teachers should develop lists of skills of both kinds; those which we expect to develop indirectly and those we must be prepared to teach directly. Categories such as: solving problems, designing experiments, collecting data, recording and processing data, interpreting data, and communicating results and conclusions can be useful to help identify specific skills.

For example, recording and processing data may require students to be able to: select an appropriate form for recording data; record data accurately in tabulated form; use simple mathematical operations to re-organize data; interpolate and extrapolate from graphs; select appropriate method of displaying data; and use charts, graphs, tables and so on to display data. Teachers could work together in groups to prepare such checklists.

Reporting assessments to students

The final phase of classroom assessment is the reporting phase, where the evaluation of the

students' performance, against the established criteria for acceptable performance is made known to those interested in it. The form of report should vary according to who is to act on it, since no evaluation should be an end in itself, but rather give evidence concerning the selection of a course of future action.

Generally, decisions about student progress might be considered as being made by either the teacher, the student, or the parent. The teacher's influence is profound. But decisions are made by other groups as well and the whole matter of decision-making has been taken up by Welch (Chapter 3). Assessment reports may be oral (most common) or in writing, and may be directed to the individual student, small groups of students, or class groups. Reports may be useful to the teacher, other teachers, employers and parents, or as feedback to those who designed the curriculum. The report phase is the culmination of assessment and needs as much care and consideration as the development of any instrument on which the report may be based.

How does the teacher of integrated science cope with all these demands for assessment? Often there is enough difficulty in simply keeping up with the content of the course. The intention has been in this paper to show that assessment is a natural and integral part of the teaching process and must be treated by the teacher as such. The best teachers may be those whose intuitive assessment ability in a class situation is most highly developed. The first thing teachers are encouraged to do is to recognize when they are making assessments in the classroom and how these assessments may lead to better decisions. One sure aid is to turn the teacher's intuitive implicit criteria into explicit criteria, agreed on by both student and teacher.

The science teacher in the integrated science class may be described as first a planner, second an organizer and third a leader. These three processes are likely to be ineffective if the teacher does not have an effective monitoring system concerning what is going on in the classroom. It is assessment that is the core of the monitoring process and must guide future planning, organizing and teaching.

Too often we think of assessment as being test-generated numbers. Numbers are important. We do need to compare our students. But we ought to explore ways of putting numbers to assessments other than pencil-and-paper tests, so that the numbers which we use to determine the future of students relate to the whole spectrum of aims and not only those which can be so assessed.

Teachers often seem unwilling to involve each other in the instructional process. Mutual sharing of ideas by one's peers can be a most helpful process. An idea crystallized by Wrigley[1] of a 'critical friend' to discuss student assessments is one such idea. A person who will 'talk through' as an equal what is proposed by the teacher for the class gives the teacher an opportunity to reflect on what is intended. This is 'reflective evaluation' in which a few minutes of discussion with a critical friend may help clarify ideas or interpret incidents which otherwise might be lost.

The role of science teacher associations and other similar organizations of peers in providing a forum for teachers to discuss their ideas on assessment should be emphasized. Teachers in their own educational context are the people in the best position to decide what may be achieved in their own local area, and it is through the teachers themselves that change will occur. There are definite limitations to what individual teachers may achieve in their classrooms without the benefit of pooled ideas. This is particularly true when teachers explore assessment outside the usual paper-and-pencil kind.

In conclusion, I wish to make some points about assessment in general. Those who are in schools which are still, or recently have been, dominated by public examinations will realize how impersonal such examinations are. They have their place, but one tendency which they have is to make classroom assessment impersonal as well. Classroom assessment is a very personal matter to the students and to the teacher - particularly the formative assessment and the assessment of attitudes and process skills.

[1] J. Wrigley, 'Fools and Angels', Schools Council, Dialogue, Vol. 15, 1973, p. 8-9.

Integrated science is a rubric for organizing science ideas in an educational way. It is educational because it involves students in learning through the interactions of student with student and student with teacher. The process of assessment is a process of discourse, of information giving and information receiving, of drawing inferences and of making decisions about people. It is one of the most sensitive activities in which a teacher will be involved. For this reason, a teacher should not make assessments too instrument-centred. Rather, they should be person-centred, with as much emphasis upon the students being able to make their own evaluative judgements against mutually-agreed criteria as upon the judgements being made by the teacher.

BIBLIOGRAPHY

Ahmann, J.S. and Glock, M.D.: Evaluating pupil growth. Fifth edition. Boston, Allyn and Bacon, 1975.

Australian Science Teachers Journal. Theme 'Testing and Evaluation'. Vol. 19, No. 2, June 1973.

Australian Science Teachers Project: Guide for teachers. Melbourne, Victorian Government Printer, 1974.

Bellack, A.A., et al: The language of the classroom. New York, Teachers College Press, 1967.

Block, J.H. (ed): Mastery learning: theory and practice. New York, Holt, Rinehard and Winston, 1971.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. (eds): Handbook on formative and summative evaluation. New York, McGraw Hill, 1971.

C.S.E.: 'An experiment in the oral examining of chemistry', Schools Council Examinations Bulletin. London, Evans/Methuen, No. 21.

Education Department of South Australia: Do-it-yourself curriculum guide for junior secondary science. Trial edition. Secondary Science Curriculum Committee, 1974.

Eggleston, J.F.: A critical review of assessment procedures in secondary school science. University of Leicester, School of Education, 1965.

Fraser, B.J.: Tests of enquiry skills. Books 1-3. Monash University, 1974.

Gray, J.H.: 'Assessment as a source of feedback to teacher and pupil', in Sutton, C.R. and Haysom, J.T.: The art of the science teacher. U.K., McGraw Hill, 1974.

Gronlund, N.E.: Preparing criterion referenced tests for classroom instruction. New York, Macmillan, 1973.

Hedges, W.D.: Testing and evaluation for the sciences. Belmont, Wadsworth, 1966.

Hudson, B. (ed): Assessment techniques. London, Methuen, 1973.

Klopfer, L.E.: 'Evaluation of learning in science', Chapter 18 of Bloom, B.S., Hastings, J.T. and Madaus, G.F. (eds): Handbook on formative and summative evaluation. New York, McGraw Hill, 1971.

Lewis, D.G.: Assessment in education. London, University of London Press, 1974.

Mackay, L.D.: 'Roles for evaluation in science courses', The Australian Science Teachers Journal, May 1974.

Mayer, V.J.: Unpublished instruments in science education: a handbook. Columbus, Ohio, ERIC Information Analysis Centre for Science, Mathematics and Environmental Education, 1974.

Nedelsky, K.: Science teaching and testing. New York, Harcourt, Brace and World, 1965.

Nelson, C.H.: Improving objective tests in science. Washington, D.C., National Science Teachers Association, 1967.

Postman, N. and Weingartner, C.: The school.book. Particularly chapter 7, 'How to evaluate a school and what to do about a bad one'. New York, Delacorte Press, 1973.

Rosenshine, B.: Teaching behaviours and student achievement. International Association for the Evaluation of Educational Achievement, 1970. Mimeographed.

Simon, A. and Boyer, G. (eds): 'Mirrors for behaviour', Research for better schools. Philadelphia, 1967.

Theobald, J.H.: Classroom testing: principles and practice. Melbourne, Longman Hall, 1974.

Whitfield, R.C.: 'Curriculum objectives, examinations and curriculum change'. Cambridge Journal of Education, No. 5, Easter 1972.

Wrigley, J. 'Fools and Angels', Schools Council, Dialogue, Vol. 15, 1973, p. 8-9.

8 Evaluating student progress in integrated science—public examinations

John F. Kerr
University of Leicester, United Kingdom

SUMMARY

We shall be concerned in this chapter with ways of improving public examinations, not only to discriminate between individuals but also to contribute to the mastery of those skills and concepts which the teacher and curriculum developer desires. The effects of public examinations on teaching and learning integrated science will be analysed, techniques of examining will be reviewed, problems identified and the way ahead considered.

Although the form of public examinations is a perennial topic of discussion, their role in evaluation tends to be disregarded. Previous chapters have stressed the need for evaluation in integrated science teaching to improve curriculum decisions for course developers, teachers and students. Traditionally, public examinations attempt to measure individual attainment or achievement of students, a process which still implies selection or sorting for something, however egalitarian the particular society may be. Since their origin in Imperial China before the beginning of the Christian era, the social intention of examination systems has been for greater equality of opportunity and less sponsorship or nepotism. Where the trend towards more open secondary and higher education systems is strong, as in North America, the selection function of public examinations is reduced; but in most countries, although competitive examinations have been severely criticised for centuries, they continue to dominate their educational systems, in practice if not in theory.

It is surprising that this aspect of evaluation - the measurement of student progress by public examinations - should have received so little attention in the appraisal of new curricula and their diffusion into the school systems. In recent times, as indicated by some papers in this volume, the emphasis has turned towards the analysis of curriculum materials and the use of observational schedules to identify styles of teaching and patterns of interaction between teachers and students. Through these new techniques and insights, our control of curriculum change is becoming more deliberate and effective. There is also strong evidence[1] that the form of public examinations is one of the most significant influences upon this complex process of curriculum change. This being so, persistence of the view that examinations are ends in themselves, neglecting their wider influence, should be challenged.

The growth of integrated science courses during the past decade or so[2], for reasons and circumstances which vary from country to country, has provided an opportunity to depart from those examination practices which were traditional with the separate sciences. The changes in

[1] S.S. Dunn (ed), Public Examinations, Australia, Rigby Ltd., 1973, p. 206; and W.B. Elley and I.D. Livingstone, External Examinations and Internal Assessments, New Zealand, NZCER, 1972, Ch. 3.

[2] Richmond, New Trends ..., Vol. II, op. cit.

practice depend, among other factors, on past history, degree of selectivity in the secondary and post-secondary systems, the ability of science teachers to cope with broadly-based science courses and the competence of the examiners. Clearly, no set of circumstances in one country fit exactly those in another, and no prescription for measuring students' progress can be independent of circumstances. The huge empirical study of the International Association for the Evaluation of Educational Achievement (IEA) on science education in nineteen countries[1] demonstrates a range of practices. Experiments with forms of school-based assessments which may be feasible in some countries may be premature in others. But there are some criteria and techniques for assessing performance which are widely applicable, irrespective of country and subject area (see Ramsey, Chapter 7).

This chapter provides a review of some of the influences of public examinations, their effects upon teaching practices, points to some related problems, and considers the way ahead.

The effects of public examinations on teaching and learning

A recent view that teaching and examining should be considered together[2] (e.g. Wrigley), represents a significant shift from a long-standing obsession with the reliability of examinations. Because criticism has been focussed on the unreliability of examinations[3], this has been for a long time the main target for reform with consequent improvements in examining techniques. Unfortunately, these developments, desirable as they are, are seen by many teachers as involving complex mathematical abstractions comprehensible only to the highly numerate. For this reason, improvements in examining practices have not often been related to the practical problems of science teaching. One of the causes for the rapid decline of general science over twenty years ago was that the public examinations did not reflect the intended approach to the teaching of these courses. Most of the examination questions set required the recall of factual information from the separate fields of biology or chemistry or physics[4]. The examinations did not match a teaching approach which was intended to have 'its roots in the common experience of children ... and (sought) to elucidate the general principles observable in nature, without emphasizing the traditional division into specialised subjects ...'[5]. The social, economic and administrative forces for the retention of established ways of examining proved too strong to be overcome by educational ideologies. Public examinations by written essay-type papers were thought to be impartial. It was useless to argue, as Eggleston did, that 'impartiality is not justice, it may be equality of injustice'[6]. Power and prestige continued to be accorded to the examinations by

[1] L.C. Comber and J.P. Keeves, Science Education in Nineteen Countries, New York, John Wiley, 1973, Ch. 7.

[2] J. Wrigley, in Schools Council Examinations Bulletin 32, Assessment and Testing in the Secondary School, London, Evans/Methuen Educational, 1975, p. 10.

[3] R.L. Edgeworth, Essays on Professional Education, London, Johnson, 1808; P. Hartog, E.C. Rhodes and C. Burt, The Marks of Examiners, London, Macmillan, 1936; and M. Dunstan, 'Sources of Variation in Examination Marks', University New South Wales Bulletin No. 2, Australia, 1966.

[4] L. Connell and W.S. James, 'General Science Today', The School Science Review, No. 138, London, S.M.A., 1958, p. 277.

[5] Science Masters' Association, The Teaching of General Science, Part I, London, Murray, 1936, p. 30.

[6] J.F. Eggleston, A Critical Review of Assessment Procedures in Secondary School Science, United Kingdom, University of Leicester School of Education, 1965, p. 10.

universities and the professions. Teachers felt they were more accountable to their students than to society which was reluctant to hand over to the teachers control of the examinations. This reluctance is gradually being overcome as evidence accumulates that public examinations do not measure certain important qualities[1]; they are often irrelevant to the needs of students; and they are not good predictors of success in post-secondary education, still less in non-academic occupations.

How can public examinations be used to improve science teaching and raise the level of performance of students? The development of integrated science projects in many countries of the world over the past decade or so has been widely publicized. Experience with many of these new courses shows how student endeavour can be stimulated and performance raised. So often, however, the public examinations set do not measure the new outcomes and abilities gained, and the results are not an accurate reflection of the student's real performance during the course. In the sections which follow, some suggestions will be made for bringing examining techniques into line with the purpose and practice of the new integrated science courses.

Techniques of examining

Although the process of course development in integrated science projects generally follows the familiar research, development and diffusion pattern (R.D. and D.) (one which is particularly suited to the teaching of science), the relationships between ways of examining and objectives of the course, and ways of examining and teaching methods, are too frequently neglected. Until the importance is realized of the matching of ways of examining with precisely-defined course objectives and also with teaching methods, public examinations will continue to be unsatisfactory in that they will measure only part of what they purport to measure.

In this chapter, only a basic outline can be given of how to improve public examination procedures. The writer believes that the information needed is in fact not difficult to understand and act upon, and that all teachers are capable of handling the elementary techniques given the willingness to modify their own outlooks and attitudes. For more sophisticated treatment of the techniques, technical works on testing are widely available, reference to examples of which will be given.

The credibility of marks or grades resulting from any examining procedure depends solely on two properties - their reliability and validity. Absolute reliability (defined as 'consistency') is impossible to achieve because of three broad sources of error due to variations in marking, restricted sampling of syllabus content or of those abilities which might reasonably be expected of candidates, and inconsistency of performance by candidates. Some of the ways by which these sources of error might be investigated and reduced are dealt with later. More detailed methods of analysis of scores to estimate the degree of unreliability will be found in specialised works[2]. The second property affecting the credibility of examination marks is validity, that is the degree to which the test does what it is meant to do. In the public examination of student progress, the problem of validity becomes that of deciding if the abilities required to answer examination questions are consistent with the objectives of the integrated science course preceding the examination. Only when course objectives are specifically defined in terms of what the student is expected to be able to do is it possible to design questions which make demands consistent with them. Empirical procedures are described in the specialized texts referred to above to determine how far validity has been established.

The primary purpose, therefore, of the rest of this chapter is to discuss ways of raising the reliability and validity of the results of public examinations in integrated science.

1 Elley and Livingstone, op. cit., ch. 4.

2 e.g. Schools Council Examinations Bulletin 3, *The Certificate of Secondary Education: an Introduction to some Examining Techniques*, London, H.M.S.O., 1964; and E.J. Furst, *Constructing Evaluation Instruments*, New York, McKay, 1961.

No prescription for measuring students' progress can be independent of national circumstances but there are some criteria and techniques which are universally applicable. The matching of ways of examining to course objectives, it has been argued, is such a principle. It can be demonstrated in generalized form by means of the matrix in Table 12.

Table 12. The relationship between ways of examining and aims in integrated science courses.

Aims	Ways of examining: Multiple-choice (objective) test 1	Essay-type written paper 2	Course-work assessment 3	Practical test 4	Project 5	Oral test 6
1. Recall of basic information	●					
2. Understanding of key concepts	●	●				
3. Higher cognitive processes (e.g. thinking critically, solving problems)		●	●		●	
4. Mastery of practical skills				●	●	
5. Ability to design experiments					●	
6. Scientific attitudes applied to Society			●			●

Clearly, this table of specifications is much simplified. For a particular public examination, it would be desirable to state the aims in more specific form before the term 'objectives' would be justified but Table 12 illustrates the possibility of using a cluster of six ways of examining, each of which is suited to measure how far certain aims have been achieved. Brief comments on the advantages, but not the disadvantages, of each method of examining follow together with examples of test-items designed to measure the degree of competence in particular skills or attitudes.

Multiple-choice tests are particularly convenient for testing recall of facts and understanding of concepts although test items can be constructed by experts to measure powers of inference and reasoning. In simple form, they are widely used as part of the examination because the answers can be marked easily, quickly and reliably, and a wide range of the course can be tested in a short time.

Example 1 tests recall of facts.

The stages in the life history of the housefly are in order,

(a) larva-egg-pupa-adult.
(b) pupa-larva-egg-adult.
(c) pupa-egg-larva-adult.
(d) egg-larva-adult-pupa.
(e) egg-larva-pupa-adult.

Example 2 tests understanding of principle (induction).

When a current is induced by the relative motion of a conductor and a magnetic field, the direction of the induced current is such as to establish a magnetic field that opposes the motion. This principle is illustrated by,

(a) a magnet attracting a nail,
(b) an electric generator or dynamo,
(c) the motion of a compass needle,
(d) an electric doorbell.

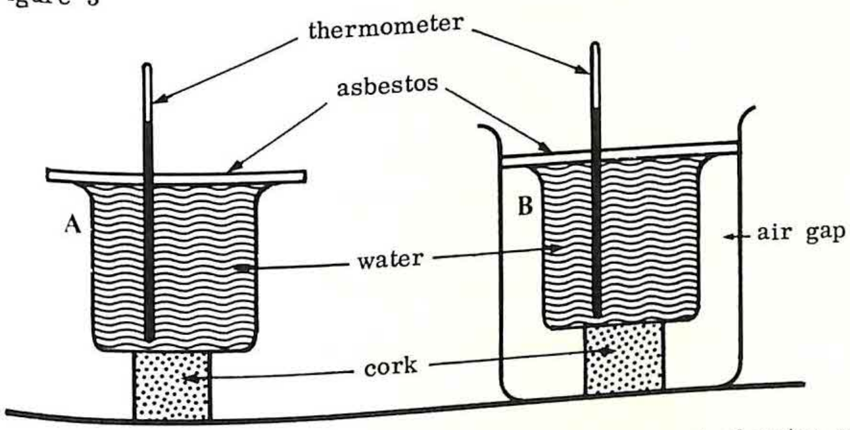
(Examples 1 and 2 from Schools Council Examinations Bulletin 4.)

Essay-type written papers continue to be commonly used, often in conjunction with objective tests. The questions set may be in a variety of forms - structured (guided) questions to test different levels of ability in the different sub-questions; questions which require comprehension of data provided and the drawing of conclusions from it; and, less frequently, free-response of open-ended questions. Although too many questions in these papers continue to test no more than basic information[1], they are used to best advantage for testing higher levels of cognitive ability and qualities such as power of organization and ability to write lucidly.

Example 3 tests higher cognitive processes (structured question).

In order to investigate the effect of an air gap on the loss of heat from a sample of water a pupil set up the experiment shown (Figure 5).

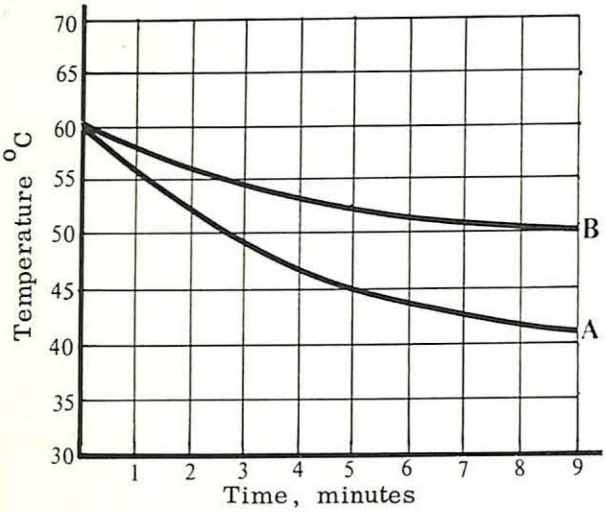
Figure 5



Two identical beakers were filled with the same amount of water at $60^{\circ}C$. The temperature of each was recorded every minute and the results were plotted on a graph.

[1] J.F. Eggleston and J.F. Kerr, Studies in Assessment, London, E.U.P., 1969, p. 25.

Figure 6



1. How does the air gap affect the loss of heat? Is it increased, decreased, or the same?

2. In experiment A, what was the temperature after three minutes?

3. How many minutes did it take for the temperature to drop 15°C in beaker A?

4. What was the difference in temperature between the water in A and B after eight minutes?

5. Which of the following materials, if placed in the air gap, would reduce the heat loss more? Underline the correct answers.

Vermiculite	Glass wool	Copper
Foamed polystyrene	Water	Rubber

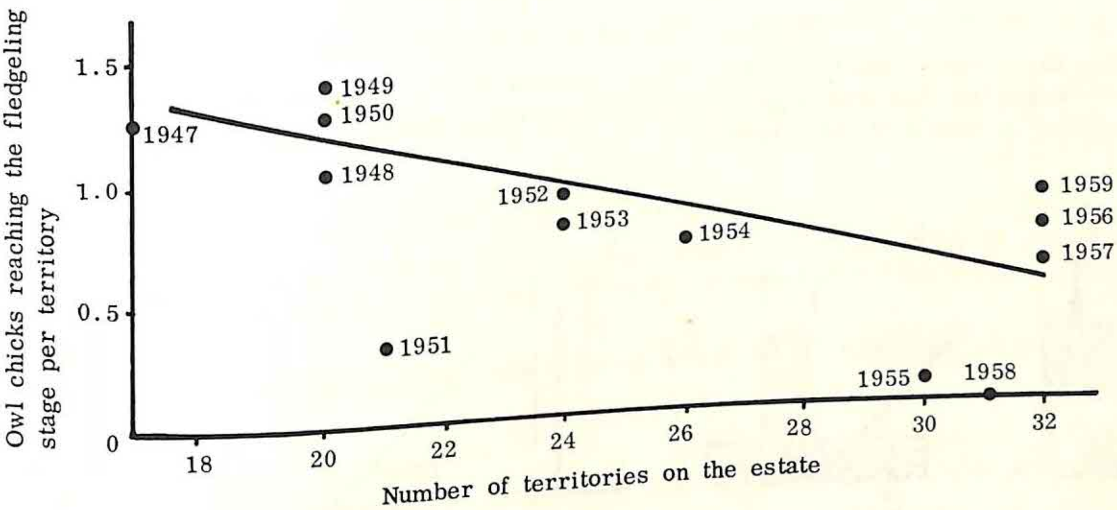
6. Explain how knowledge of the results of the above experiment would enable a house-builder to reduce heat losses in a house.

(Example 3 from Nuffield Secondary Science, Examining at C.S.E. level, London, Longman, 1972.)

Example 4 tests comprehension of data and drawing of conclusions.

Like many other birds, pairs of owls establish a 'territory' in the breeding season. They patrol the territory and from it obtain food for themselves and their young chicks. Other owls rarely trespass. The data below shows, over a period of years, the number of territories on a particular estate, each one of which is held by a pair of owls, and the average number of young birds reaching the fledgeling stage per territory.

Figure 7



(a) Describe how the number of chicks per territory reaching the fledgeling stage changes with time.

(b) Describe how the population of owls changes with time.

(c) Suggest how the change in population with time leads to an explanation of the changes in the number of chicks reaching the fledgeling stage.

(Example 4 from Schools Council Integrated Science Project (SCISP) Paper 5, 1972.)

Course-work assessment needs careful planning and is most effective when selected assignments from a variety of different types of activity on the course are specified for inclusion in the assessment. It provides a strong incentive to a student to keep up with work. It offers immediate and repeated feedback. It reduces the peak of examination pressure for teacher and student at the end of the course. It is not an alternative to tests but a measure of some of the less tangible skills and attitudes acquired. Systematic record-keeping of profiles of grades awarded is necessary.

Practical tests are intended to measure the degree of mastery of practical skills which are not shared with theoretical work. The attention given to individual practical work varies greatly. In the I.E.A. study only two out of the nineteen countries completed the optional practical tests[1]. The evidence suggested that practical tests measure abilities quite different from traditional tests.

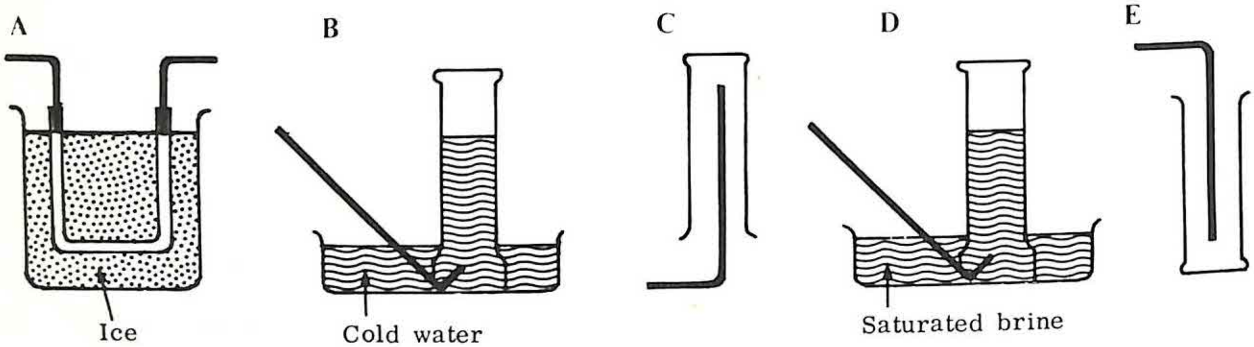
[1] Comber and Keeves, op. cit., p. 29.

and from tests designed to measure practical skills without access to apparatus. (Professor Sim used sets of colour transparencies in a study in Malaysia, 1974.) If the unique value of science in the curriculum arises from acquiring knowledge through experiment, observation and problem-solving, the skills involved should be precisely specified and evaluated[1]. A form of internal assessment closely related to the practical work carried out on the course is generally preferred to external practical examinations, which have a restricting effect on the range of laboratory work carried out in schools and usually base the marks awarded on the results only. If an assessment of the practical ability of candidates in public science examinations is not attempted, experimental work may be neglected on the course. Five methods of assessing practical work are described by Macintosh[2].

Example 5 is a multiple-choice test item intended to measure an aspect of practical ability.

A substance has a boiling point of 21°C. In its gaseous state it is denser than air and soluble in water. From the diagrams below, select the one representing the most suitable apparatus by which to collect a sample of the substance.

Figure 8



(Example 5 from Comber and Keeves, op. cit., p. 377.)

Projects provide an opportunity for study in depth and usually enhance motivation. They give scope to show originality in the design of experiments or field work and to demonstrate practical skills. They can be very time-consuming for both teacher and student but they measure a wide range of abilities. The topic for the project should be precisely stated, feasible and cover a limited problem for the practical study of which resources are available.

Example 6 tests ability to design experiments and practical skill.

Devise two ways to measure the speed of a model racing car and compare the results.

[1] E.g. J.F. Kerr, Practical Work in School Science, United Kingdom, Leicester University Press, 1963, p. 21; and H.G. Macintosh (ed), Techniques and Problems of Assessment, London, Arnold, 1974, p. 98.

[2] Macintosh, op. cit., p. 90.

Oral tests are unreliable, often not feasible, time-consuming and favour extrovert students. They assess verbal fluency, self-confidence and can reveal scientific attitudes and reasoning powers not otherwise apparent. Although the tradition of oral examining continues in some countries (e.g. in eastern and western Europe), particularly in relation to practical ability, it is not normally recommended by the writer at this level.

To provide a more specific example of the task of relating an examination set in several parts to a detailed specification of objectives, the revised version of the SCISP pattern is reproduced below. This[1] course has been referred to frequently in this volume of New Trends in Integrated Science teaching as well as in Volume II. It is a course of integrated science of double O-level value for 13-16 year-old pupils. The standard provides a satisfactory basis for A-level science courses in the United Kingdom.

AIMS

It is hoped that pupils will achieve the following aims:

Skills

Pupils should be able to demonstrate their degree of competence in

1. (a) recalling, and (b) understanding those concepts which would enable them to pursue science (courses in Physics, Biology, Chemistry of Physical Sciences) to a higher level or as a hobby;
2. (a) recalling, and (b) understanding those patterns which are of importance to the scientist;
3. making critical appraisal of available information (from whatever source) as an aid to the formulation or extraction of patterns;
4. using patterns and making critical appraisal of available information (a) in order to solve problems, and (b) make reasoned judgements;
5. organizing and formulating ideas in order to communicate them to others;
6. understanding the significance, including the limitations, of science in relation to technical, social and economic development;
7. being accurate in the reporting of scientific work;
8. designing and performing simple experiments (in the laboratory and elsewhere) (a) to solve specific problems, and (b) to show perseverance in these and other learning activities.

Attitudes

Pupils should:

9. be able to work (a) individually, and (b) as part of a group;
10. (a) be sceptical about suggested patterns, yet (b) willing to search for and to test for patterns;
11. be concerned for the application of scientific knowledge within the community.

The content of the work is based on three fundamental ideas - building blocks, interactions and energy. Throughout the three years there will be a continuous search for 'patterns' which are important generalizations and these will then be used to solve problems of a practical and a theoretical nature. The project aims to educate through science as well as in science, and so is particularly concerned with the sociological implications and technical applications of science.

[1] Schools Council Integrated Science Project (SCISP), Patterns: Teachers' Handbook, London, Longman, 1975.

Ideally it is hoped that a single teacher will eventually be able to deal with the course. There are possibilities for team teaching, but it would, of course, be contrary to the nature of integration if the material were divided into subject areas separately taught by specialists.

WAYS OF EXAMINING

Integrated Science leads to a double credit at O-level. A single credit is available. At present, five papers are set, together with teacher assessment.

Paper 1	Essay/short answer paper testing aims 4b, 5, 6.
Paper 2	Essay/short answer paper testing aims 4a and b, 5, 6, 10a.
Paper 3	Multiple-choice paper testing aims 1a and b, 2a and b, 3.
Paper 4	Multiple-choice paper testing aims 1b, 2b, 4a.
Paper 5	Essay/short answer paper testing aims 4a and b, 5, 8a.

Teacher assessment testing aims 5, 7, 8a and b, 9a and b, 10b and 11.

One of the commonest criticisms of public examinations is their undesirable backwash effect on methods of teaching. It is important that the techniques of examining should be compatible with the teaching methods used to achieve particular purposes. For example, there is little possibility of the spread of investigatory methods of teaching which lead students to find out for themselves if the examination demands mainly recall of facts. The possibility of matching ways of examining to teaching styles can be illustrated by means of a matrix.

Table 13. The relationship between ways of examining and teaching methods in integrated science courses.

		Ways of examining					
Teaching methods		Multiple-choice (objective) test	Essay-type written paper	Course-work assessment	Practical test	Project	Oral test
1.	Instruction/teacher-directed, didactic	●	●				
2.	Practical demonstration/teacher-direct, participation	●	●				
3.	Practical work/teacher-directed, finding out				●		
4.	Project work/student-directed, finding out					●	
5.	Discussion/collaboration, theoretical						●
6.	Solving problems/teacher-directed, theoretical			●			
7.	Private study/any of above			●			

Since different teaching styles are likely to result in the development of different abilities, an examination pattern which encourages a teacher to use a variety of teaching methods is desirable. As far as science teaching is concerned, one of the studies carried out to evaluate teaching styles

in relation to student scores on attainment and attitude tests was completed by Galton, Eggleston and Jones[1] with the co-operation of ninety-six teachers and 3,000 students. Although the findings are written in tentative terms, there is evidence that even with new science courses which emphasize guided-discovery methods, many teachers do not question students about their observations during practical work or encourage them to hypothesize or design experiments. The conclusions suggest a lack of match between the curriculum developers' aims and practice in the laboratory. The need for a close inter-relationship between the three components - examinations, aims and teaching methods - should never be overlooked.

An international view of examination patterns in integrated science, including all types of combined science courses, reveals a rapidly changing scene and a growing awareness of the basic techniques to which reference has been made in this chapter within a wide variety of practices. A great deal of the integrated science examining is at Junior Certificate level (as in Kenya) after only two or three years science. A common pattern is that which is found in the United Kingdom, east and west Africa, the Caribbean countries, Malaysia, Hong Kong, Singapore, Australia and New Zealand. It comprises two or more structured papers, usually including some objective questions, together with a form of practical assessment. A trend in many of these countries is to replace external examinations by internal assessment. Although there is evidence of the wider use of different types of questions, little attempt is being made to assess higher, cognitive skills, for example by testing the ability to handle data or draw conclusions (as in Examples 3 and 4 above). A number of countries, including the United Kingdom India, the United States, France and Sweden are using internal assessment as a component of the total examination grade awarded or for comparison with (or moderating) the externally-awarded grade. On the other hand, in most central European countries and elsewhere where grading is by internal assessment only, the process is being widely criticised because of its unreliability, variations in standards between schools and influence on pupil/teacher relations. The pooling of national resources in west Africa and east Africa as well as in the Caribbean to form regional examination boards has resulted in the setting of questions more relevant to local circumstances. The schools in another group of countries, including the United States, Canada and Japan, use an internal system of credit awards which are often checked by externally-prepared standardized tests. A trend in the socialist countries of eastern Europe, China and Cuba is towards the assessment of groups of students, a pattern which is said to lead to the development of self-assessment and a sense of responsibility[2].

In the following section, some of the factors are mentioned which need to be taken into account when these conflicting practices are compared and improvements in public examinations in integrated science are considered.

Problems of examining integrated science

So far in this chapter our concern has been about increasing the validity of public examinations. The problems of implementing the proposals are mostly about reliability and arise from organizational and technical limitations in the examining process.

The most obvious difficulty is the lack of expertise of science teachers as examiners. This is more serious in a decentralized system which has a higher measure of internal assessment than in a closed centralized system which usually works through an examinations board. The importance of training teachers in examining methods is dealt with later. Another organizational

[1] M.J. Galton, J.F. Eggleston and M. Jones, Processes and Products of Science Teaching, United Kingdom, Schools Council Research Series (Macmillan), 1976.

[2] See R.E. Lister's paper on 'Trends in Techniques and Criteria Used in Assessing Student Achievement in Biology Education' at the International Congress on the Improvement of Biology Education (UNESCO), 1975, for further information about examinations in the countries mentioned above.

problem is ensuring that adequate information is readily available to students as well as to teachers about the examining system, particularly about the nature and purpose of each part of the examination. It often leads to unnecessary apprehension and tension if students are not aware of the skills which each part of the examination is intended to measure. It is desirable that more than one examiner should mark scripts independently (except for objective tests) and candidates should not sit more than one examination per day. This is not usually possible because of cost and timetabling difficulties.

A number of common fallacies arise from a lack of appreciation of the 'numbers game' played in public examinations. The total mark obtained by pooling scores in a number of tests each of which measures different abilities (for example, by adding marks from course work, a project and a practical test) is almost meaningless. A profile of grades is more informative. Errors in the reliability of marks are caused by differences in the standard of marking between examiners and in the spread of marks over the mark scale. The continued prestige in some quarters of the percentage mark - 40 per cent is a pass, 70 per cent a distinction - is a strange bit of fiction. The application of simple statistical methods can help to minimize many of these errors[1].

No more than a brief reference is possible to other standard methods of improving reliability which are a little more technical, some of which involve the application of statistical techniques. One of the major problems with school-based examinations is to achieve comparability between the marks from different schools. Relative achievement levels between schools must be reliably known if the grades are to be acceptable for the award of certificates or selection purposes. Methods of moderation are available to determine the standards of achievement reached by pupils in a particular school relative to those of other schools[2]. In particular circumstances (for example, in the different regions of Malawi where educational opportunities available to students vary from one region to another), each region is allocated a different achievement level for comparability purposes.

Another problem area which influences the validity as well as the reliability of examinations is the complex matter of sampling. The sampling of skills and teaching methods has already been mentioned. The sampling of course content is considered by Ramsey (Chapter 7). When a choice of questions is offered on examination papers, it is equivalent to setting a number of quite different papers, often a very large number of them. If a candidate is asked to answer five out of ten questions on two papers, this is equivalent to offering over 60,000 question papers. It is also important to ensure that a test-item, particularly with objective tests, is not too difficult or too easy (the facility value, F) and that it discriminates between good and poor candidates (the discrimination value, D). This technique of item analysis is described simply in Examinations Bulletin 3[3]. The problem of measuring the affective outcomes of an integrated science course has not been mentioned. "Attitudes to science" scales are available[4] but are not generally used in public examinations. The assumption is that gains in scientific interests and attitudes as a result of the course will be reflected in the candidate's performance in the examinations.

The way ahead

There can be no definitive answer to the problem of how to measure student progress in science or any other subject. A system found acceptable in one country because of its stage of economic

[1] See, for example, Schools Council Examinations Bulletin 3, op. cit., and Macintosh, op. cit., Ch. 8 and 12.

[2] These methods are described fully in Elley and Livingstone, op. cit., Ch. 8-15.

[3] Schools Council Examinations Bulletin 3, op. cit., p. 51.

[4] N.F.E.R., 1973.

development and educational history may be unsuitable for another country. While public examinations in the United Kingdom at present are becoming more open, giving more responsibility to teachers, some European countries are discarding course assessment because the skill of teachers to control quality and reliability is in doubt. Since the methods used reflect the values and assumptions of existing society - though usually after a substantial lag in time - there is a need to keep examinations continually under review. It should be the responsibility of a named group to initiate discussion about the reform of examinations on the basis of community needs; for example, by a body such as the Schools Council for the Curriculum and Examinations established in England in 1964. This body has pioneered the reform of examinations in integrated science through SCISP to encourage a much wider range of outcomes and it is now financing the after-care of the project to facilitate diffusion of the new methods. The Caribbean integrated science project (CARISO) sponsored by Unesco fulfils a similar role.

Looking ahead, it seems likely that the guidance function of public examinations will increase as emphasis on the selection function decreases. The real problem is to develop in science teachers and examiners the understanding, technical competence and perceptions necessary to cope with and indeed encourage this change. If we don't, judging from past experience, examining processes will be relatively unaffected. There is an urgent need for all teachers to be familiar with the basic principles and techniques of examining mentioned in this chapter. An introduction to examination systems and the principles of assessment should be included in initial training courses for all teachers as well as forming a regular feature of the further professional training of practising teachers. Lecture courses on these matters are inadequate. Structured workshop sessions are preferred in which small groups work closely together on real problems, at the same time drawing theoretical principles from practical exercises and discussions. It is necessary for tutors responsible for these workshops to assemble a collection of materials which are relevant to local circumstances and have been shown to be effective. One of the sections of the Activities and Experiences collection[1] entitled Feedback to teacher and pupil (FTP) includes units on ways of examining, a teacher's oral questions, what is being tested in the examination, devising a good question, designing an examination paper, marking and item analysis. With appropriate modification these units, sample parts of which are reproduced below to indicate the approach, might form the basis for a workshop.

Sample 1

Students are asked to review all the items in the exhibition, make a note of each type of item, what they think it tests and where they might use it in their teaching.

The exhibition

This could be a walk-round exhibition, or alternatively items could be passed around the group. The following are possible ones to include (with pupils' responses wherever possible):

1. Essay questions.
2. Structured questions.
3. Fixed-response questions.

Sample 2

What skills are being tested in each of the following items? Categorize each according to the highest ability which you think it tests. (22 items in unit.)

1. A red powder is known to be an oxide of a metal. It dissolves in dilute nitric acid giving a pale blue solution. Which of the following could it be?

A. Ferric oxide Fe_2O_3.

B. Cuprous oxide Cu_2O.

C. Mercuric oxide HgO.

D. Red lead Pb_3O_4.

[1] Feedback to Teacher and Pupil, topic 9 in Science Teacher Education project, op. cit.

4. Questions requiring pupils to produce drawings or diagrams.
5. Short question tests.
6. A practical examination paper.
7. Various homework assignments.
8. Personal and impersonal accounts of experiments done in class.
9. Accounts of projects.
10. A short tape-recording of a question-and-answer session in class.
11. Question books, e.g. those of the Nuffield O-level Physics Project.
12. Comprehension questions about the content of a passage.

Through the Eyes of the Pupil contains material suitable for several of these categories, as also does chapter 9 of The Art of the Science Teacher. Tutors using the unit in trials included other material on assessment in the exhibition, e.g. booklets from examination boards, complete GCE and CSE papers and published collections of questions so that students know where to get material. The exhibition would be broadened by adding intelligence tests, creativity tests, attitude tests, etc.

Discussion, and briefing for private study

A closing plenary session provides opportunity for the students to ask questions about the materials in the exhibition. They can then be asked to devise either a short question test or a longer written test on a topic of their choice, applying what they have learned in this session.

E. Zinc oxide ZnO.

2. The construction of the evolutionary family tree for reptiles and mammals is based upon

A. Degrees of similarity and differences between the structure of their bodies.
B. Evidence from blood tests.
C. The evidence of fossils only.
D. The knowledge that in evolution small species are replaced by larger species.
E. Experiments with inheritance.

3. Man's food resources would decrease if

A. Green plants became more numerous.
B. Herbivorous animals eaten by man used energy more efficiently than at present.
C. The food chains between plants and man became shorter.
D. The numbers of steps in the food chains between plants and man increased.
E. Light penetrated deeper into the ocean than at present.

4. When the speed of a car is doubled its kinetic energy

A. Remains the same.
B. Is $\sqrt{2}$ times as great.
C. Is twice as great.
D. Is four times as great.
E. Is increased in proportion to the acceleration.

5. In order to launch a rocket to the moon, it is proposed to carry out the launch from a satellite in space. The main advantage of this method is that

A. The distance to the moon is less.
B. The gravitational attraction of the moon is greater.
C. Less accuracy is needed in firing the rocket.
D. The thrust needed for the rocket is less.
E. Loading operations are simpler.

Sample 3

Guidelines for the criticism and improvement of test items.

A. Structured questions, built from short answer items, to each of which the candidate constructs his own response.

1. Does the stem describe unambiguously and economically a situation that is intelligible and interesting to the candidates?
2. Are the items following it short, unambiguous and answerable in a sentence or two?
3. Do they test worthwhile skills?
4. Is there a gradient of difficulty from the early items to the later ones?

B. Fixed-response items in general (but more particularly, multiple-choice items).

1. Is the bulk of the information in the stem, so that the options are short?
2. Do the options follow fluently and grammatically as extensions of the stem?
3. Are the options plausible?
4. Does the item discourage guessing, and encourage the type of thought it is hoped to test?
5. Can the correct option be identified in any other way besides reasoning as intended?
6. Does the item test a worthwhile skill?
7. In a series of items, can a candidate who gets early items wrong nevertheless still score on later ones?

Sample 4

Marking by impression and with a mark scheme. (Copies of about five scripts supplied to each student.)

First reading

Read through the scripts quickly, and award each a mark out of 20 by 'impression', i.e. by a general reading forming an overall judgement. Do not award marks for specific points at this stage. Enter the marks on your mark list.

Sorting method

Now try a different method. Look at each script and decide whether it is average, below average, or above average, and place it on one of three piles accordingly. Then take the pile of average scripts and spread these scripts in front of you in rank order. Treat the top and bottom piles similarly. Examine carefully the borderlines between the piles, and make adjustments as necessary. Award the middle script 10 marks, and the other scripts 0-20 as you think appropriate. The extremes of the mark range need not be used. Enter the marks and position in class on your mark sheet. Now compare your ratings with those of your colleagues.

Criteria

Now make a list of the criteria you used in awarding higher marks to some papers than others.

Devising a mark scheme

Compile a mark scheme for the question, using a total of 20 marks. The marks allocated for each point should be specified.

Using it

Mark the script that comes first in alphabetical order, using your scheme, and then compare your scheme and the mark awarded with those of your colleagues.

Finally, you may like to agree on a mark scheme with your colleagues and then remark the whole batch. Compare the marks with those awarded by impression. Is it worth having a detailed scheme?

Other matters calling for attention in the immediate future are the question of record-keeping, particularly to control intermittent assessment of course and practical work; the monitoring of

laboratory work and its relation to theory; question banking[1], that is the collection of a library of test items of known statistical characteristics for use by examination boards and teachers; and more understanding of the non-cognitive aspects of learning science. Examining the broad spectrum of student progress in science learning could be so much more than measuring what is known. Research and development work on examinations has already provided useful guidance, some of which is referred to in the following bibliography. Owing to growing public demand for accountability in education, the conflict of role between the testing and guidance functions of examinations seems likely to continue. It is to be hoped that reform will be guided more by co-ordination between examination practices and educational enquiry than by political expediency.

BIBLIOGRAPHY

Lauwerys, J.A. and Scanlon, D.C. (eds), 1969, The World Year Book of Education 1969: Examinations, London, Evans Brothers.

Scottish Education Department, 1969, Science for General Education, Curriculum Paper No. 7, Edinburgh, H.M.S.O.

Sutton, C.R. and Haysom, J.J. (eds), 1974, Assessment as a Source of Feedback to Teacher and Pupil (J.H. Gray), The Art of the Science Teacher, England, McGraw-Hill.

[1] Macintosh, op. cit., Ch. 13.

9 The role of evaluation in adapting learning experiences in integrated science to differences between pupils

Wynne Harlen
University of Reading, United Kingdom

SUMMARY

This chapter begins with a discussion of the importance of adapting learning experiences to differences between pupils. What this means in practice is described in terms of a model which represents the adaptation of the learning environment as a succession of adjustments, a process in which evaluation plays a central part. A series of dimensions is defined describing variation in features of the learning environment which are throught most relevant to catering for differences between individuals. Various programmes, at primary, middle and secondary level, are outlined and described by profiles using these dimensions. Finally, some of the ways which teachers can use to gather information about the individual characteristics of pupils which affect learning are discussed and illustrated with examples from programmes developed in different countries.

INTRODUCTION

Wherever decisions are being made, there is a role for evaluation to play. Among the many different kinds of decisions which teachers of integrated science face daily in their work is an important set which concerns the learning experiences of individual pupils. Within any class or group, the pupils vary to some extent in such things as interest, level of development of concepts and cognitive and manipulative skills, attitudes, previous experience, general intelligence and rate of learning. Course materials vary in the manner and degree to which they allow for variation between pupils; some have flexibility and alternative activities built into them, in others teachers may select, devise or adapt activities to suit the individual characteristics of their pupils, and in some there is very little provision for accommodating to differences between pupils within a class. Any attempt to provide learning experiences which vary to suit pupils involves taking decisions based on information about the pupils. The gathering and using of information about individual pupils in this context is the kind of evaluation which is the concern of this chapter.

The role of evaluation considered here differs from that discussed in other chapters in that it is part of a teaching-learning strategy for integrated science teaching, built into the strategy in acknowledgement of the value of providing pupils with experiences appropriate to various aspects of their development and to other characteristics which affect their learning. It is for this reason that we begin with a brief discussion of the rationale for the importance of catering for individual differences, followed by a suggested model of the way it is carried out in practice, showing the role which evaluation plays. After this, various approaches to the teaching of integrated science through methods which take account of individual differences will be considered, taking examples from many different countries. In the final section, we draw together various methods which can be used for evaluating individuals as part of a strategy for helping each pupil along a unique path of development.

The importance of individual differences

One of the strongest and most obvious reasons for the use of teaching methods which take variations between individuals into account is that such variations are important in determining achievement. Variations which arise from differences in home background, measured intelligence and other aptitudes, have a far greater influence upon learning[1] than teaching methods and school environments. In a later chapter of this book, Brown shows how very little of the variation in attitudes to science of thirteen-year-old pupils was accounted for by the characteristics of the science course or class, whilst about half was ascribed to characteristics of the pupils which were present when the pupils entered school. Brown's results are typical of others[2] which have shown that differences such as those already mentioned and others which could be added - motivation for learning, ability to learn through various modes, personality factors such as reflectivity or impulsivity[3] - influence learning to an extent which tends to swamp differences in teaching methods. When teaching ignores differences in these things and provides the same conditions and experiences for all pupils, those whose learning style and development happen to be better provided for will have more opportunity to learn than those less well matched by the activities available. The latter pupils may be equally capable of learning the content, given different conditions (e.g. a longer time for working, more concrete manipulation, larger or smaller steps, or an opportunity to encounter the same concepts in a different context). In consequence, when the teaching does not bend itself to fit the characteristics which add up to a person's 'aptitude' for learning, the differences in aptitude are reflected in performance. On the other hand, if learning experiences are varied to match individual characteristics to a greater degree, more pupils will have opportunity for further learning, each in the manner which suits, and starting from the point so far reached.

Another reason emerges from the changing emphasis on different goals of education; science teaching, it is urged, must be concerned, in addition to cognitive goals, with the development of responsibility, self-discipline, decision making and openmindedness in the students[4]. For these goals, "processes by which new problems are met are more relevant than answers from the past[5]". The goals are more likely to be achieved through methods which require students to take responsibility for their learning, to be self-monitoring and to find things out for themselves, than through conventional methods. This applies to "separate science" teaching as much as to integrated science teaching. However, though the goals of integrated science teaching do not differ radically, there is a conviction among its proponents that integrated science enables process and attitudinal goals to be achieved more effectively (see Mayer, chapter 5). Achieving these goals requires teachers to find out about their pupils as part of the teaching-learning process, for if a teacher aims to help a pupil solve problems or see patterns in observations, it is necessary to know how the pupil views the problem or something about ability to interpret observations. Integrated science provides the teacher with a longer time of contact with pupils than separate science and therefore provides greater opportunity for teachers to know their pupils, in compensation for creating a demand for teachers to use this opportunity for gathering and using information about variations between pupils.

[1] B.S. Bloom, Stability and Change in Human Characteristics, New York, Wiley, 1964.

[2] Summarized by H.J. Walberg, "Models for Optimizing and Individualizing School Learning", in K. Marjoribanks (ed) Environments for Learning, Windsor, Berks, NFER Pub. Co. Ltd., 1974.

[3] J. Kagan, "Reflection - impulsivity; the generality and dynamics of conceptual tempo", Journal of Abnormal Psychology, Vol. 71, 1966, p. 17-24.

[4] M.D. Andrew, "Schools, Science and Society", Science Education, Vol. 54, No. 4, 1970, p. 319-324; and L.E. Klopfer, "Science Education in 1991", School Review, Vol. 77, No.3/4, 1969, p. 319-324.

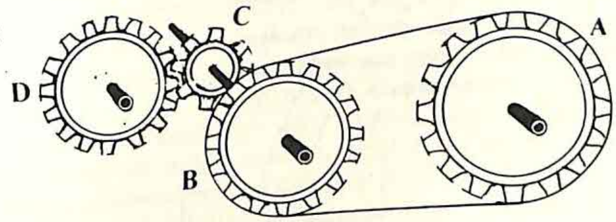
[5] V.M. Howes, Individualization of Instruction, New York, Macmillan, 1970, p. 5.

Adapting activities to pupils in practice

It is important to be clear about what is meant by adapting learning experiences to individuals as well as to say what is not meant. What it is taken to mean here is that the range of experiences for which the individual learns is such that demands are made at the right level to encourage progress. Since each pupil is unique, it is unlikely that identical sets of experiences are equally profitable to all, and thus there will be differences, sometimes large, sometimes only slight, between the activities of pupils. What this individual catering does not mean, is that the pupils will be working alone. Partly because catering for individuals has sometimes been implemented through programmed instruction and teaching machines, there is a common tendency to associate 'individualized' learning with 'isolated' learning. As defined and interpreted here, adapting to individuals is not related to the separation of students during the process of their learning nor does it depend on the use of educational technology. In fact, no assumptions are made about class organization beyond the essential provision of flexibility and opportunity for the teacher to have individual contact with each pupil; this could be within the context of large or small group working, and does not preclude occasional whole class discussions nor private study. An example is probably most helpful in explaining the meaning.

A class of twelve-year-olds was working with cogwheels, looking at the effect on speed and direction rotation of fixing cogs of various sizes together, sometimes touching and sometimes connected by a chain. One group had set up an assembly like this:

Figure 9



The teacher asked them to find how many times D rotated for every one rotation of C. As they were using the handle at B to turn the cogs one of the group, John, said A was not necessary any more. Robert disagreed saying "That one makes all the others go round". The teacher suggested Robert should investigate the effect of taking the chain away and look carefully at how each cog was connected to the others. Meanwhile John, who seemed quite clear about the cause and effect relationship involved, was working out how to mark how far D rotated for one turn of C. The third member in the group quickly saw that it was not necessary to move the cogs to find the answer and she was busy counting the teeth on C and D. It was clear to the teacher that Robert's notion of cause and effect in this situation was at a level such that he would not have benefited from measuring the rotation of the cogs; it would probably have been a meaningless exercise for him. But also the girl in the group had been able to grasp the relationship involved more clearly than John, and she would have not benefited from his approach.

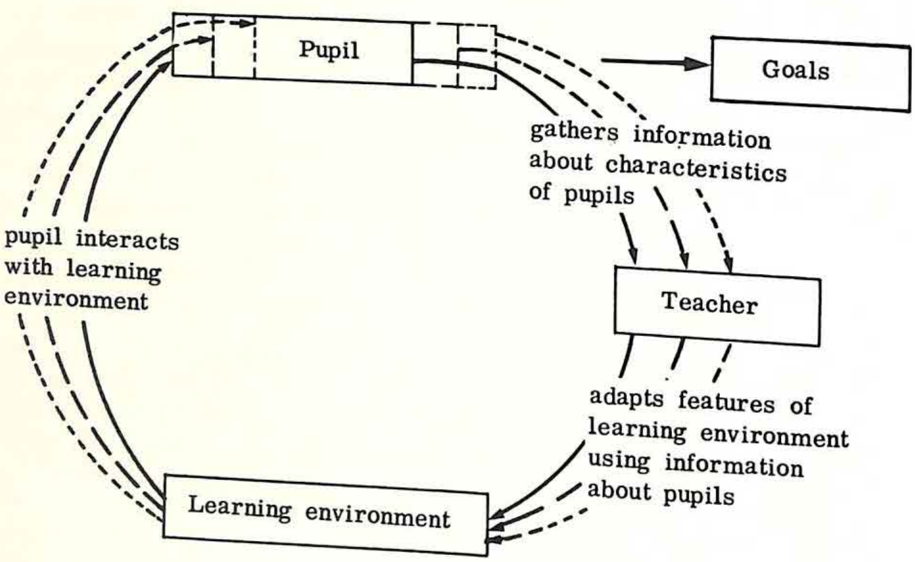
When decisions are made about pupil activities, there are many factors to take into account. The most obvious are the ones relating to the level of difficulty, complexity and elaboration of the possible activities. These have to be taken into account when considering the objectives which the pupil might achieve. Then there are various ways in which any particular idea can be grasped: through experiment, through discussion, through reading, watching film or television. There is generally also a variety of contexts in which the ideas could be presented and which could be chosen according to the interests and experience of the pupils. But this is not all; there are other variables concerning supervision, encouragement and incentives. Together, these make up some of the factors in the learning environment of the pupils.

There are problems, however, when it comes to making decisions as to how to adapt various aspects of the learning environment to the pupils. It has to be acknowledged that there is a serious lack of consistent evidence from research to answer questions about how to promote

learning in pupils with different characteristics. Walberg[1] reviews the results of research which has attempted to relate treatment to aptitude for learning, and concludes that very little that is conclusive has come out of the considerable effort so far expended in this area. It seems unrealistic to expect that a set of guidelines will emerge for adapting activities to any particular combination of abilities and learning characteristics.

In the absence of prescriptions for the most effective course of action in particular circumstances, a practical approach is to adapt the learning environment to individual pupils through an iterative process. The notion of 'matching' experiences to pupils is seen as one of making successive adjustments to the ever-changing situation. The process can be visualized as the repeated cycling through a feedback loop. In this loop, the gathering of data about pupil characteristics is followed by its use in adapting certain aspects of the learning environment with a view to enabling pupils to make progress towards the goals of their learning. After each cycle of the loop, the effect of the adaptation is included in the information taken in about the pupil in the next cycle. As the process of attempting to match by using the information goes on (that is, after each cycle), the pupil is some way further towards achieving the goals.

Figure 10 Matching as a succession of cycles



As with any model of a complex process, there are cautions to keep in mind. The attempt to analyse the process into its parts inevitably means that these appear as distinct and separate in a way which does not happen in the real situation. It may be impossible to distinguish a cycle of feedback when a teacher is moving from one pupil to another and is making a number of decisions very closely together and she may be unconscious of the steps in the process. Nevertheless, given the time to rationalize decisions, she could justify them by reference to her knowledge of the pupils and her appreciation of the goals of their work. It is also necessary to consider the loop represented above as one element in a model for a whole class. Since the adaptation of the environment is carried out to suit each child, the complete model would have as many elements of this kind in it as there are pupils in the class. Further, we should not be so optimistic as to suppose that the only changes in the pupils are towards the intended goals. Some unintended (and perhaps unwanted) outcomes will emerge, but it is the purpose of the complex process of 'matching', represented by the model, to reduce as far as possible the more damaging outcomes which arise from boredom or frustration when there is serious mismatching.

[1] Walberg, op. cit.

The role of evaluation

In the strategy which has just been described, evaluation plays a central role. It provides the information about a pupil's present ideas, knowledge and skills which is essential for making decisions as to how to help make the next steps. An essential feature of this continuing evaluation is that it is a starting point for action, not an endpoint or a summary of the effect of actions taken.

Taken individually, some of the decisions may seem rather small and trivial: for instance, to give direct help to this pupil, but let that one work it out independently; to give a working sheet to one group but encourage another group to plan their own investigation; it is the sum total of small decisions like these which creates the learning environment for the pupils. Teachers have, of course, to make decisions anyway about the features of the learning environment; what distinguishes teaching strategies which take account of individual differences from those which do not is that in the former the decisions are taken after gathering and taking into account information about the pupils as individuals. By contrast, in the other approaches, though the teacher may have information about individuals (at least as far as their performance is concerned) from marking assignments, setting tests, question and answer sessions, the pupils are nonetheless treated as a homogeneous group and one set of decisions applies to all.

Approaches to catering for individual differences in integrated science teaching

As the practical problems of implementing a teaching strategy which adapts experiences to pupils are very great, it is no surprise to find that different approaches have been taken in attempting to surmount them. Like pupils, teachers have individual characteristics and adapt methods to their own abilities and preferences. The constraints caused by external pressures and classroom conditions (e.g. class size, the age of the pupils and total years of attendance at school) are among the factors which make one approach more suitable than others in particular circumstances. We shall try to distinguish between approaches to adapting experiences to individuals according to the features of the learning environment which are regarded as fixed or varied for each pupil.

There are many aspects of the teaching-learning process which can be varied. Gibbons[1] has produced a comprehensive list of dimensions - for example, 'attendance', 'pace of study', 'teaching focus'. To cater for individual differences, he described each of these within four levels of freedom. However, not all of these are under the direct influence of a teacher's decisions and it is perhaps best to confine the present discussion to key features of the learning environment which appear to affect learning by individuals and are subject to decisions taken by a teacher.

Given that pupils vary (see Ohashi, chapter 16, for an example of the wide variation in children's responses to the same problem in attitudes, interests, level of development of concepts and cognitive and manipulative skills, style of learning, previous knowledge and experience, general intelligence and rate of learning), what are the elements of the learning environment which might (at least in theory) be adjusted so that individuals have optimum opportunity to learn? A list of the elements which might be affected by the teacher and could be varied - in some but not all circumstances - would include the following:

(a) Time allowed for learning - can be the same for all the class, or vary between groups or vary between individuals.

(b) Content, the activities and materials - can vary between groups or between individuals according to the interests and previous experience of the pupils, or can be kept the same for the whole class.

(c) Pupils' participation in decisions about the learning environment - selection of activities, sequence, time and goals of learning can be wholly controlled by the teacher, carried out by teacher and pupils together or left largely for the pupils to decide.

[1] M. Gibbons, What is Individualized Instruction?, <u>Interchange</u>, Vol. 1, no. 2, 1971, p. 28-49.

(d) Variations in expectations and goals - whether the content of activities varies or not, what the pupils are expected to achieve from them can be the same for all, or different for groups or individuals.

(e) Class organization - can be such that the unit is the whole class, or groups or individuals.

(f) Influence of diagnostic feedback - feedback about pupils can be an important basis for decisions about pupils' experiences or there can be no feedback gathered or used.

(g) Teacher-pupil relationship - there can be a marked hierarchy in this relationship, with pupils expected to show respect and acquiescence to teachers, or there can be a non-hierarchical relationship, based on mutual respect.

(h) Pupil-pupil relationships - mutual help in the achievement of individual or shared goals can be encouraged or discouraged.

The possibilities which have been given here for each of these eight aspects are intended to indicate the range of variation along a dimension. There is no dichotomy but the possibility of very many different values along each dimension. This is made clear in Table 14 where the direction from left to right is one of increasing provision for variation between pupils.

Table 14. Provisions for individual differences.

	Variable	Range of variation		
(a)	Time allowed for learning	fixed at same amount for all pupils	extra time available for slower to catch up	entirely flexible, according to needs of the individual
(b)	Content of activities	same activities carried out by all pupils	basic common core for all with remedial or enrichment for some	varies according to individuals
(c)	Pupils' participation in decisions	pupils take no part in choosing what they do	pupils and teacher jointly make decisions by agreement	pupils free to choose what they do and how
(d)	Variation in expectations and goals	expectations are same for all	minimum level same for all, more expected of some	expectations vary between individuals
(e)	Class organization	decisions relate to class as a whole	unit for decisions as a group (3 to 5)	decisions about pupils as individuals (within groups)
(f)	Influence of diagnostic feedback	little feedback gathered or used	can modify but not determine course of action	important basis for determining course of action
(g)	Teacher-pupil relationship	hierarchical, pupils expected to acquiesce	teacher respects pupils' opinions but makes decisions	non-hierarchical, mutual respect as of equals
(h)	Pupil-pupil relationship	pupils discouraged from discussion with each other	free to discuss but no expectation of mutual help	pupils encouraged to discuss and help each other

The list might well be extended, but serves to show the variety of elements which affect

learning but over which a teacher has control. In some circumstances, a teacher's freedom of action may be restricted by constraints set within or from outside the school, but generally these operate to reduce the range of variation which is possible, rather than entirely remove freedom. In most situations, teachers have to make some decisions about the range of items which have been listed. If there is in practice no variation between the learning experiences of individual pupils, it may well be that this is because such variation is thought either to be unimportant, impracticable or undesirable.

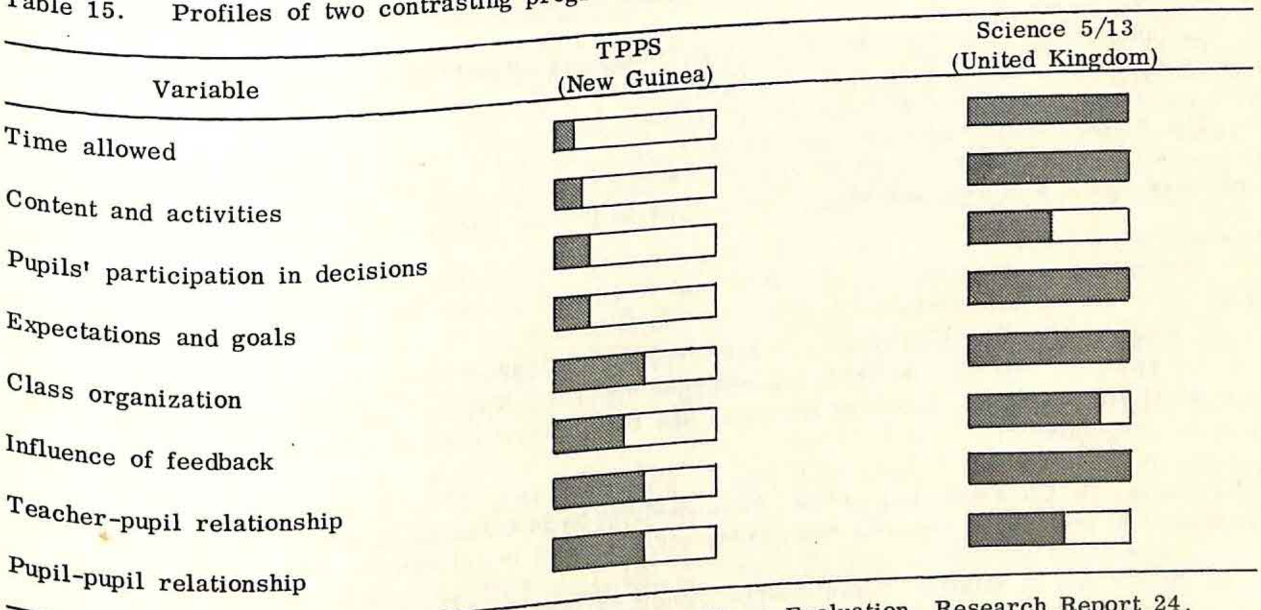
These variables can be used to describe the characteristics or different integrated science programmes and to reveal the extent to which a programme allows for variations between pupils. Two examples serve to show how programmes can be described in terms of positions along these dimensions. They are somewhat contrasting programmes for primary schools:

> "The T.P.P.S. course is one of pupil activity. Each weekly 45 minute science lesson ... is described for teachers on a card ... . The card gives the title of the lesson, the class organization (usually groups of four), a list of materials required, brief instructions for the activity, a picture of children engaged in the activity and sometimes questions, further information, recording procedures and further additional activities". [1]

> "(Science 5/13) held that teachers were capable of being, and should be, responsible for thinking out and guiding the activities of their pupils. To maintain teachers' freedom and give them help, was considered important, so that activities could follow and form the interests, and match the intellectual development of the children. ... the units do not constitute a course; they are seen as sources of ideas and guidance which teachers can use in planning and carrying out programmes of work which they devise to suit the unique requirements of their own classes". [2]

These extracts are not enough to give all the data needed for describing the programmes in terms of the dimensions, so further information has been used in drawing up the following profiles of the two programmes. The shaded areas in Table 15 show the degree to which the variation between pupils is catered for by various features of the programmes:

Table 15. Profiles of two contrasting programmes.



[1] M. Wilson, Three Phase Primary Science. Phase Three Evaluation, Research Report 24, University of Papua, New Guinea, 1974, p. 1.

[2] Harlen, op. cit., p. 8.

The Science 5/13 units were intended to enable teachers to provide pupils with experiences varying according to interests and development, whilst for the TPPS this was not a major intention. The profiles show a distinct difference, but also show that any one programme has variations in the degree to which its various features enable adaptation to individual pupils. We now use this framework to examine some integrated science programmes which have been designed to cater for certain differences between pupils, looking first at programmes for children up to the age of thirteen.

Primary and middle school programmes

Two United States elementary science programmes which have been developed for individualized learning are the I.S. (Individualized Science) programme[1] and the Personalized Programme in Science[2]. The first of these is based on the philosophy of individualization which was implemented in IPI (Individually Prescribed Instruction)[3]. It is a complete K-8 programme arranged in seven levels. For each level there are several 'units', some to be studied by all and some optional. The units are in turn divided into 'lessons', the first of which includes a 'placement test' which a student will mark independently, and use the results to plan, with the help of a 'Planning Booklet', which other lessons of the unit should be studied. For each concept, the programme provides various kinds of learning resources - e.g. individual lessons, directed group activity, reading, explorations - and the student decides which to use; after discussion with the teacher, this is recorded in the 'Planning Booklet'. There will be certain science content lessons which are taken in a given order, and assessment of concept attainment is built into these lessons. After each Check Up, an assessment sheet suggests the student's next activity, based on the score, so guidance is obtained from the materials in deciding the route from lesson to lesson. At the end of a unit, the student takes a post-test and on the basis of this the student and teacher decide whether the unit has been satisfactorily completed. Thus it is true to say that, because of the large number of options built into the materials, individual students may follow unique paths and take a part in the decisions about what the path is to be. However, student choice of diet is limited, as it were, to ready-made dishes; their own fare is not created from the raw ingredients. The programme provides for the learning environment to be adapted to the characteristics of individuals in so far as it contains suitable elements. Beyond this, the student has to adapt to what the programme provides.

In Table 16 the main features of this programme are summarized in terms of the dimensions. Again it should be pointed out that the justification for all the positions cannot be given fully here, but is contained in the descriptions of the programme[4].

The Personalized Programme (developed by Linn, Chen and Thier) is designed for minimum dependence upon the teacher; indeed this programme could be used in situations where the adult is not necessarily a trained teacher. The aim is to provide pupils with experiences matched to

[1] L.E. Klopfer, "Individualized Science: relevance for the 1970s", Science Education, Vol. 55, No. 4, 1971, p. 441-448; and A.B. Champagne and L.E. Klopfer, Individualized Learning in Individualized Science. Learning Research and Development Centre, University of Pittsburgh, U.S.A. (Unpublished).

[2] M.C. Linn, B. Chen and H.D. Thier, "Personalization in Science: a pilot study". Paper presented at meeting of National Association for Research in Science Teaching, Chicago, April 1974.

[3] J.O. Bolvin and R. Glaser, "Development aspects of Individually Prescribed Instruction", Audiovisual Instruction, 1968, Vol. 13, p. 828-831.

[4] Champagne and Klopfer, op. cit.

their intellectual level and the approach chosen to implement this is to "allow children to choose their own experiences within a general framework of possible choices. One assumes that children, in this atmosphere, will choose activities which will help them to learn"[1]. In the development of the programme it was found, however, that pupils not used to devising their activities tended, among other things, to carry out only superficial investigations with no follow-up on preliminary findings. It was therefore decided to provide the pupils with guidance as to what to investigate and how to start on it, given a particular set of equipment. The guidance takes the form of written 'challenges', posing a problem and giving suggestions for starting on it; further hints are given on the paper if the pupil gets stuck. The activities are designed to be carried out by one pupil, though they can be performed in groups. Students are encouraged to complete a challenge and write a report of their work before selecting their next one. The only restriction in their choice is that no more than one of each set of equipment is used at a time in the class.

As the summary profile of this programme shows, on-going evaluation is effectively absent. Instead, it is the pupil who makes decisions about the next actions, except in so far as there are certain rules and obligations, for example, a report has to be written after each challenge. It is argued that the pupil is in the ideal position to make the decisions; after all, no-one knows better how able the pupil feels to tackle a new problem, what is really interesting, etc. This is true, but the decisions should also be taken with the goals of learning in mind, and in the knowledge of the potential opportunities offered by different courses of action for achieving these goals. It is doubtful whether pupils have the conspectus necessary to make optimum choices and there is a real risk that progress will not be made if the decisions are left entirely to the pupils. Therefore, against the advantages of pupils having first-hand access to knowledge about themselves, one must weigh the disadvantage of their ignorance about the over-all goals and various routes to achieving them.

The African Primary Science Project (APSP) is in marked contrast to the programmes so far considered. In the latter, decisions by the teacher do not play a large part; the teacher has a prominent role, since "APSP is not just a list of scientific content to be taught to children; it is more fundamentally a way of working with children".[2] The project materials comprise a collection of units for teachers. Teachers choose which units to use, when to use them and in what sequence. The teachers' guides urge that pupils be allowed to work freely with materials, but the teacher's role in developing thinking about what is found and in introducing new activities is clear. "The teacher can help his pupils test their explanations. He can help them make simple apparatus. He can also introduce other activities which the children have not thought about, but which may help them gain a better understanding of the problem they are trying to solve".[3] It is implicit in this statement that the teacher has to evaluate the situation and to decide when to make interventions - the programme will not tell. Unfortunately, neither does the programme give any guidelines in making these important decisions; the process by which activities can be adapted to individuals is not spelled out, and the decisions are left to intuition.

The teacher's role in the Nuffield Junior Science Project is even more central, the material being quite unstructured. Considerable stress is laid in this project on matching children's interests and stages of development; specific experiences are not proposed except as examples or starting points and the pupils devise their own activities. In other words, pupil choice is encouraged, but in this case it is choice not from a prepared range of activities but from all the

[1] M. Linn, 'Individualizing Science: How do children learn to interpret experiments?' (mimeo) available from the author at Lawrence Hall of Science, Berkeley, California.

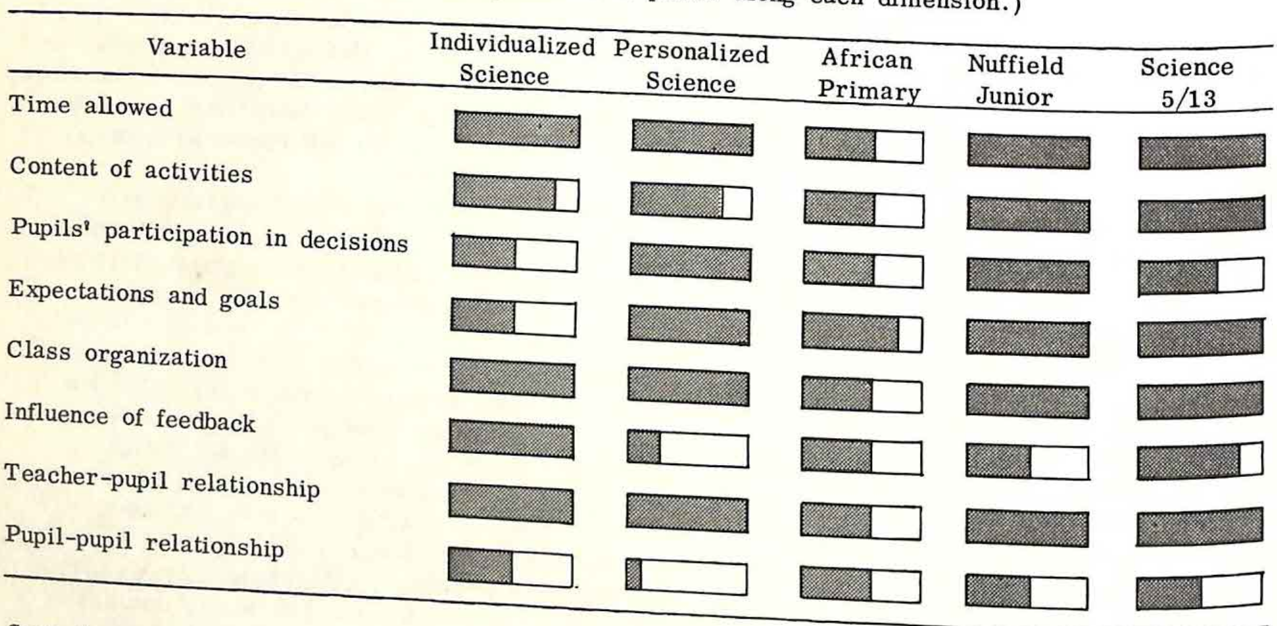
[2] E.A. Yoloye, Evaluation for Innovation. APSP Evaluation Report, University of Ibadan, Nigeria, 1971.

[3] African Primary Science Project, Teachers' Guide: 'Inks and Papers' as reproduced from the trial edition in Richmond, New Trends ... Vol. II, op. cit.

activities thought of which are at all practicable. The same doubts apply here as expressed earlier concerning whether pupils can effectively not only match but also stretch themselves in this free choice situation. Moreover, the Nuffield project emphasizes that "Once the children are interested in and handling things it is imperative for the teacher to discuss with them"[1]. The guidance during discussion has, of course, to be appropriate to the individual child; but, as in the case of APSP, no help is given to the teacher with this difficult but crucial part of teacher role.

Science 5/13 also depends on teachers to make decisions about appropriate objectives and activities for their pupils. In this project, it is intended that goals, ways of working, time for learning and interactions of all kinds should be set by reference to the pupils' characteristics. The project provided help in this matter by giving general guidelines as to children's stages of development (using Piaget's framework) and by suggesting activities and objectives suitable at three different stages. Evidence showed that for many teachers, broad indications of activities suitable for pupils at different points in development were not sufficient. This led to the establishment of a project whose concern is specifically to help teachers with the process of providing learning environments suited to individuals[2].

Table 16. Summary of features of programmes for pupils up to age 13.
(Please refer to Table 14 for the interpretation of points along each dimension.)



Secondary programmes

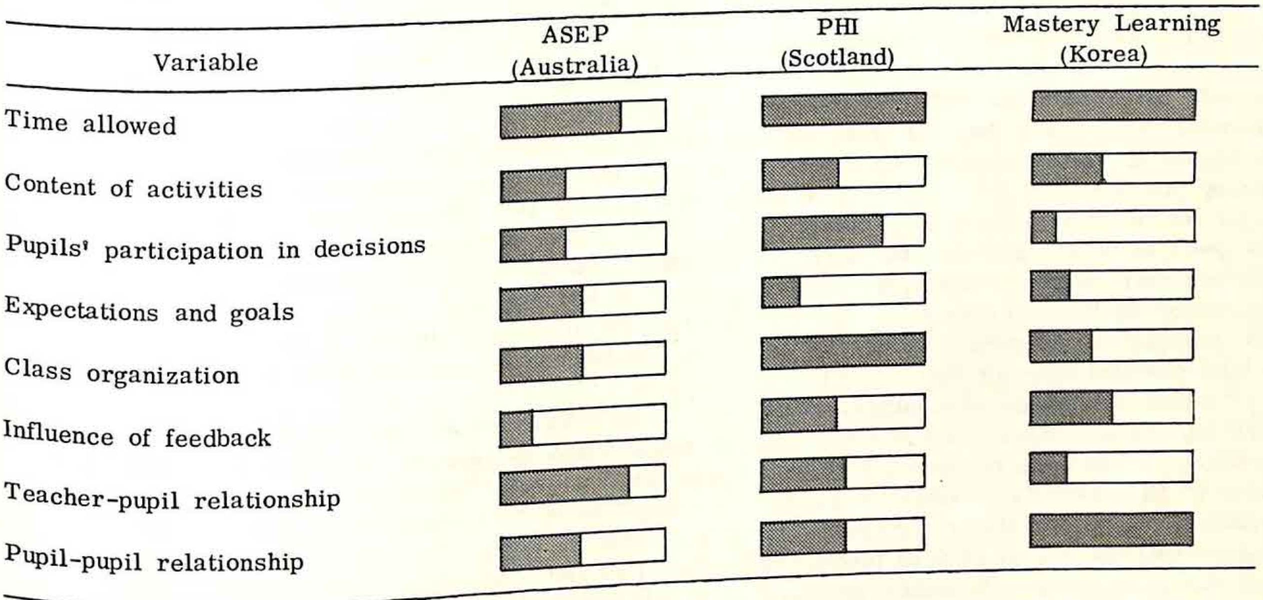
At the secondary level, there are generally more constraints creating practical difficulties to catering for individuals than at the primary level. The end of school examinations, more rigid timetabling, teaching by science subject specialists (often in separate physics, chemistry and biology laboratories) these are some of the factors which tend to operate against the use of flexible integrated science programmes. The arguments for the importance of adapting to individuals are no less cogent, but the circumstances demand a compromise between what is desirable and what is practicable. Three programmes chosen to explain approaches to catering for individual differences

[1] Nuffield Junior Science Project, Teachers Guide I, London, Collins, 1967.

[2] Progress in Learning Science, Check-lists for Earlier Development, Draft edition, 1976.

are the Australian Science Education Project (ASEP), a programmed adaptation (PHI) of the Scottish Integrated Science Project, and the Korean Mastery Learning Project in Integrated Science. The features of these programmes are summarized in Table 17 in terms of the same descriptions as used for the primary and middle school programmes. A comparison of the profiles in Table 17 with those in Table 16 shows that individual variations are accommodated to a smaller degree in all of the secondary projects than in the ones for earlier years.

Table 17. Summary of features of secondary programmes.
(Please refer to Table 14 for the interpretation of points along each dimension.)



The Australian ASEP project is a science programme for 13 to 16 year-olds designed to allow for individual differences among students. The forty-one units for pupils, with associated materials and handbook, do not constitute a course, but are separate sections of work, each taking about 10-15 hours of class time; they can be combined in various ways by a teacher to make up his or her own course. "A typical unit begins with a section (known as the core) which is completed by all students and is planned to occupy about one-quarter of the time a student spends on the unit. The actual time spent varies with individuals, as students are expected to attain a minimum level of competence. The remainder of the time is spent on options, which form most of the rest of the unit. Students are invited to choose the options that interest them. Some options are intended to extend the intellectually capable while others provide less intellectually demanding or more concrete activities" [1].

There are diagnostic tests in most ASEP units which are self-administered and marked by each student. However, these tests come at the end of a section, generally only at the end of the core; they serve to "reinforce the instructional material", but do not supply continuous feedback which could be used to adapt activities during the course of the work. It is acknowledged by the programme authors that these pencil-and-paper type tests do not evaluate achievement of non-cognitive objectives and teachers are urged to do this by using other methods, which are mainly to be found by reference to other sources.

[1] Australian Science Education Project, A Guide to ASEP, 1974, p. 10.

The Scottish Project PHI[1] (Programmed materials for the Highlands and Islands), aimed to adapt sections of the Scottish Integrated Science Scheme (see Jeffrey, chapter 12) and to produce supplementary aids which would enrich the learning environments of pupils in isolated and small secondary schools. The project explored the value of programmed materials in compensating for inadequate laboratory provision and understaffing. The programmed materials produced had the following features:

"1. They impose a measure of structure on the learning situation.
2. They require active participation by pupils.
3. They incorporate feedback on the pupils' expected responses or sensations"[2].

The pupils work individually through a work book which gives step-by-step instructions and information. At intervals, the pupils take tests based on the content which should just have been learned; they mark the tests themselves. . At these points, or at any other time when they encounter difficulties, they ask their teachers for further explanation. Occasional branches are introduced in the programme, for example, to teach the use of certain equipment for those not already familiar with it. The PHI materials thus provide for different times for learning, but variations between individuals in other aspects of the learning environment are very limited. The goals to be achieved are the same for all pupils and the modes of learning, though varied in different parts of the programme, cannot be varied in any one part; interaction between pupils is determined by the programme. The tests provide an on-going evaluation of the grasp of facts and concepts, which gives immediate feedback to the teacher and to the pupils, thus helping them to take responsibility for their learning.

Whilst PHI arose as a compensating measure in schools where one of the problems was often too small a class size, a programme in Korea which bears some similarity to it was developed to meet the opposite problem of overlarge classes. The Korean Mastery Learning Project[3] had perforce to adapt the mastery strategy[4] to operate where class size is about seventy pupils. An integrated science programme for 13 to 16 year-olds has been devised which incorporates the use of various instructional strategies and three types of evaluation. The flow chart over page describes the components and their inter-connexion.

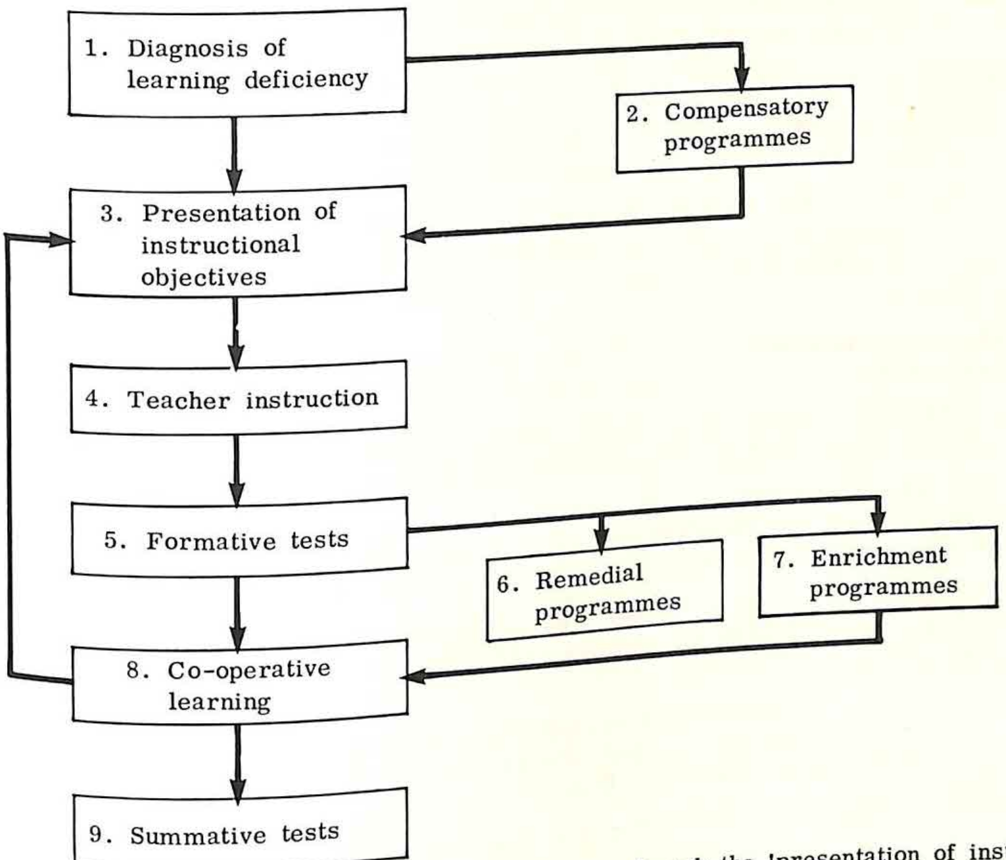
[1] M. Roebuck, J. Bloomer and D. Hamilton, Project PHI, University of Glasgow, Department of Education, 1974.

[2] J. Bloomer, Project PHI and Programmed Learning, Technical Supplement 17, Department of Education, University of Glasgow, 1974, p. 3.

[3] H. Kim, Experimentation in Educating Mastery Learning in Korea. Paper prepared for a seminar on "The Evaluation of the Qualitative Aspects of Education" held at IIEP, Paris, September/october 1974.

[4] J.B. Carroll, "A model of school learning", Teachers College Record, Vol. 64, 1963, p. 723-33.

Figure 11: Flow chart of mastery learning strategies.



Most of the components are self-explanatory, though the 'presentation of instructional objectives' and 'co-operative learning' may require enlargement. Using the teachers' guide, the teacher tells the students, before they begin each section of work, what they have to do and the standards of performance expected. The goals are the same for all the students, since it is a basic premise of the mastery strategy that "most students can achieve mastery if they are allowed and do spend the necessary amount of time on a learning task"[1]. Thus the ability to achieve certain objectives is not considered a relevant variable in this strategy. A feature which is included to help students achieve the common goals is the co-operative learning component, designed to allow students to help each other in small-group discussion with guidance from the teacher.

Kim[2] does not describe this programme as being individualized since the content and kinds of activities are not varied for different pupils, but it is the intention of the mastery learning strategy to adapt the learning environment to the individual as far as is practicable and this is done through the variation in time for learning and the use of compensatory enrichment and remedial programmes. As is seen in Table 17, the result does not compare unfavourably with PHI which claims to be individualized.

Though the programmes outlined for all school levels differ in many features, as Tables 16 and 17 show, they agree upon the importance of a pupil having learning opportunities which are

[1] B.S. Bloom, " Mastery Learning", J.H. Block (ed) Mastery Learning, New York, Holt, Rinehart and Winston, 1971.

[2] H. Kim, et al, Mastery Learning in the Middle School, final report on Mastery Learning Project, March-August 1970, Seoule, Korean Institute for Research in the Behavioural Sciences, 1970.

adapted to present capabilities as much as possible. In order to provide a pupil with such opportunities, it is necessary to know about him or her as a learner - the how, what and wherefore of learning. Various methods are available for finding out about the pupils' characteristics, and it is to the discussion of these that we now turn.

Methods for on-going evaluation of individual pupils

As a starting point for discussing the pros and cons of various methods for gathering information about pupils, we can consider what, ideally, is required of any method.

1. It should be capable of providing information about the key characteristics relevant to learning, which will include pupils' interests, attitudes, level of development of concept and cognitive and manipulative skills, style of learning, previous learning, rate of learning, motivation for learning.

2. It should enable information about relevant items from those mentioned to be gathered frequently and repeatedly.

3. It should give detailed information; anything as general as an IQ score, for instance, blurs out variations existing in one person between abilities of different kinds.

4. It should give information about where a child is with regard to some criteria of behaviour or levels of development rather than state how he rates against the standard of other children, i.e. it should be criterion-referenced rather than norm-referenced[1].

The possible methods of assessment can be envisaged as spread along a dimension from the highly structured, through the semi-structured to the unstructured. Typical characteristics at three points in this continuum are:

highly structured:	time and administration rigidly controlled, generally group tests, marked against externally set standard (examples: external exams, IQ and other standard tests).
semi-structured:	time and administration can be varied as long as performance on given task is fairly undertaken, can be group or individual, marked by teacher or pupil against given criteria (examples: teacher-made tests, diagnostic tests, either written or performance).
unstructured:	carried out at any time, generally individual, assesses performance in normal activities against behavioural criteria (examples: observation schedules and check-lists, informal interviews).

When considering the suitability of methods in different parts of this range, it is also necessary to keep two constraints in mind besides the requirements which have already been mentioned: (a) the teacher's convenience, and (b) the effect on the pupil of constant information collection. The burden on the teacher should not be so great that the task becomes irksome, thereby creating a risk that gains in feedback could be lost in commitment. Similarly, pupils should not feel that they are always under scrutiny and that they spend undue time completing tests rather than engaging in more interesting learning activities. These two constraints are difficult to reconcile, since relieving the teacher may mean using ready-made tests for the pupils. This dilemma faced the Korean Mastery Learning Project, where the class size precluded direct diagnosis and observation by the teacher and frequent tests for the pupils were introduced instead. In this case, special steps were taken to ensure that the pupils understood the point of the tests, could act on the results, and undertook the task in a co-operative rather than a competitive spirit.

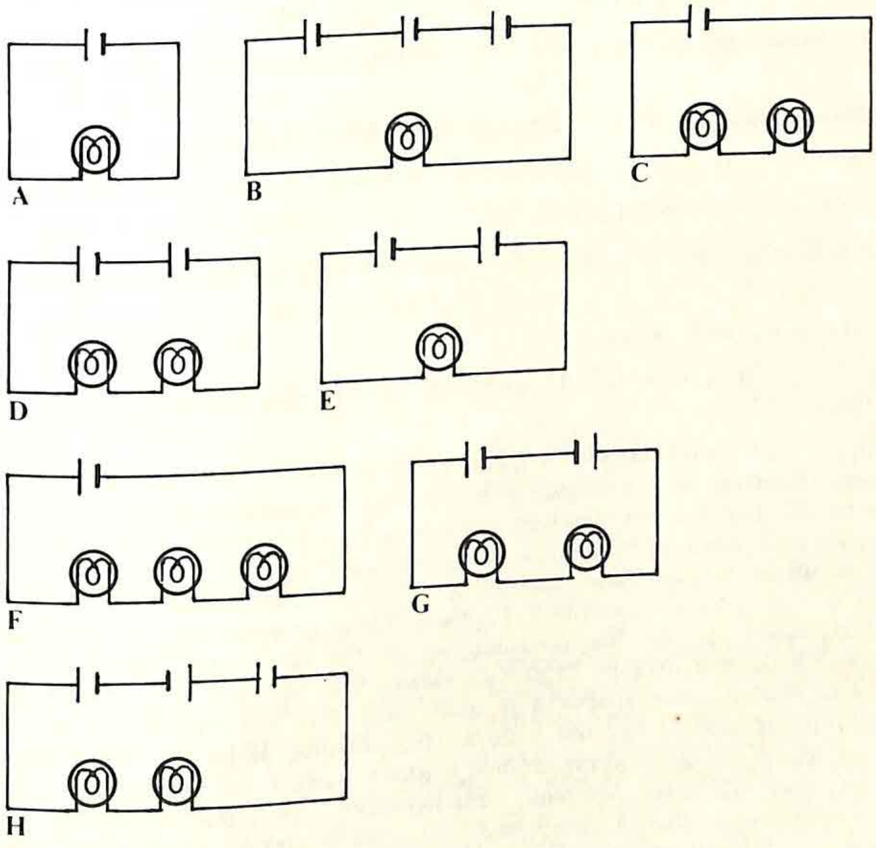
[1] R. Glaser and R.C. Cox, "Criterion-referenced testing for the measurement of educational outcomes", R.A. Weisgerber (ed), Instructional Process and Media Integration, Chicago, Rand McNally, 1969.

From the list of ideal requirements for a method of evaluation, it is clear that highly structured tests are quite unsuitable for the purpose being considered here. Semi-structured tests are more suitable; these may be either built into the programme and taken at specified points, or provided by the programme and used as desired by the teacher. The I.S. programme, PHI and Korean Mastery Learning Project all provide examples of tests built into the programme. The test item shown below comes from the PHI unit 'Electricity in Motion' and is placed just after the part of the programme where the effects of number of bulbs and numbers cells connected have been investigated separately and in combination.

Figure 12

Block F: Test Card 6

1. Here are some circuit diagrams. Look at each one, and arrange them in order according to how much current would flow. Just write the letters A to H inclusive in your jotter, starting with the one with least current, finishing with the largest current. If you think any two circuits will give the same current, underline the two circuits.



Source: PHI, Workbook 'Electricity in Motion', Project PHI, Department of Education, University of Glasgow, 1973.

As an example of tests provided for teachers to use as they choose, the 'Evaluation Supplements' of the elementary science project SCIS (United States) can be taken. The method

suggested for checking progress could be used equally well for on-going evaluation. The teacher is instructed to prepare a specified set of the evaluation materials and ask certain questions and is given criteria for judging satisfactory understanding of the points involved. This example comes from the evaluation supplement for the unit 'Organisms'.

> "Invite one or two children at a time to a quiet corner and give them time to examine the materials on the tray. Tell them you are interested in what they have learned about planting seeds and growing plants. First ask:
>
> 1. Name the materials you would need if you were going to plant these grass seeds. Do you need anything else?
>
> After each child has had an opportunity to identify the objects he would use, ask:
>
> 2. How would you use these materials? or Describe how you would plant the seeds and grow plants.
>
> Now ask the following convergent questions:
>
> 3. How deep would you plant the seeds?
> 4. Here are three bags of soil that have different amounts of water in them. Which one shows best how wet the soil should be for growing plants?
> 5. How long after planting grass seeds will it be before plants begin to grow out of the soil?
>
> Criteria. The criteria below will help you recall each child's responses:
>
> 1. Did he select all of the necessary materials for planting: seeds, soil, planter, water?
> 2. Did he describe the process of planting?
> 3. Did he recognize that the seed should be covered with soil but should be near the surface?
> 4. Did he select the damp soil sample?
> 5. Did he suggest a period of time for the grass to emerge that approximated the classroom evidence?"[1]

What are the advantages and disadvantages of these kinds of test for the purpose of on-going evaluation? Practical considerations will probably help decide whether it is feasible to use tests built into the programme or whether teacher chooses when to administer them. In either case, these tests are able to probe children's performance on a particular subject or skill in a highly detailed way. Whilst it would be far too time consuming to cover many skills or concepts at once in such detail, taken one at a time, immediate feedback about progress can be obtained. However the testing of skills one at a time has implications for the instructional programme, since it is assumed that the skills are learned in series rather than in parallel. The method is evidently unsuitable where an unstructured approach is used since then the feedback is required across a wide range of characteristics at any one time as the learning skills which will be required at the next step are not pre-determined. A disadvantage which these tests share with all others is that the choice of content may affect the outcome; furthermore, since the tests have to be short, it may not be possible to obtain more than a small sample of the behaviour which is being assessed; this does not allow for normal fluctuations in interest, concentration and performance. In addition, the method is suitable for checking skills, concepts and knowledge, but not for the other characteristics relevant to learning. For these it is necessary to turn to the methods which are at the unstructured end of the continuum outlined earlier.

[1] SCIS Organisms Evaluation Supplement 1972.

It can be argued that special situations do not always have to be set up for finding information about pupils' progress, for if the normal activities really do provide opportunity for the goals of learning to be achieved, they also provide opportunities to see how far development has reached towards these goals. In circumstances where it is practicable for observations to be made during normal learning activities, this can be the basis for a method of gathering information which has several advantages. Observation does not interfere with the pupil's work and can be carried out on different occasions, giving a variety of behaviour samples. A range of many different characteristics can be assessed by observation, and it has most of the features suggested as desirable for the purpose under discussion. On the other hand, observations are subjective and less reliable than other forms of assessment. 'Observation' in its fullest meaning includes listening and responding as well as looking, and is a skill which requires training. While it may not take up the teacher's time when it becomes a normal part of work, some training or mental preparation may well be required before this state is reached.

There is a great danger that observation provides little more than personal opinions, with low reliability and, consequently, low validity also. Therefore, if it is to be the basis of a viable method for gathering feedback about pupils, it cannot remain totally unstructured. It requires a structure which comes from considering the goals of children's learning and from developing criteria which can be applied in interpreting children's behaviours in terms of progress towards the goals.

Duckworth provided teachers using APSP with general guidelines, which are clearly related to the affective goals of the programme, to help them use their observations of children during lessons to find out whether they were making progress:

"Here are some questions a teacher can ask himself as he watches a child's work from day to day:

1. Does he make suggestions about things to do and how to do them?
2. Can he show somebody else what he has done so they can understand him?
3. Does he puzzle over a problem and keep trying to find an answer even when it is difficult?
4. Does he have his own ideas about what to do, so he doesn't keep asking you for help?
5. Does he give his opinion when he does not agree with something that has been said?
6. Is he willing to change his mind about something, in view of new evidence?

---

23. Does he ever watch something patiently for a long time?
24. Does he ever say, "That's beautiful"?"[1]

At a more detailed level, the project 'Progress in Learning Science' (United Kingdom) has developed check-lists to provide a framework for observation. Each check-list covers about twenty-five abilities, concepts and attitudes which are among the more important goals of science activities for pupils between 5 and 13 years of age. Some examples are 'observing', 'identifying variables', 'raising questions', 'communicating - verbally and non-verbally', 'interpreting findings', the concepts of 'area', 'volume', 'force', 'energy', 'adaptation', and attitudes of 'curiosity', 'openmindedness', 'responsibility', 'willingness to co-operate'. Statements in the check-list are intended to indicate typical behaviours at three levels in the development of each item.

[1] E. Duckworth, Evaluation of the African Primary Science Programme, EDC, 1970.

For example:

<u>Interpreting findings</u>:

1. Rarely goes beyond a straight report of observations showing little sign of appreciating any patterns within them which could have been discerned.
2. Makes rather sweeping statements about his observations, suggesting patterns which are not justified by the evidence.
3. He puts 'two and two together' and attempts to find patterns in his findings which are justified by what he has observed.

<u>Openmindedness</u>:

1. Tends to stick to preconceived ideas ignoring contrary evidence and behaving as if unaware of the existence of opinions or findings different from his own.
2. Changes from one idea or opinion to another inconsistently being influenced by the authority behind them rather than the force of the evidence or argument.
3. Generally listens to and considers all points of view; accepts ideas different from his own if the evidence is convincing.[1]

The check-lists are intended to structure observation, focusing it upon aspects of behaviour deemed to be significant in relation to the goals of science education in the early years. They are used constantly in this way as a basis for guiding teachers' decisions without any record necessarily being made of the observations. But they can also be the basis for a cumulative record of development which shows up areas of progress or lack of it and can guide longer term decisions about providing children with experiences which build upon children's strengths to remedy their weaknesses.

In the past few years a variety of methods have been devised for finding out pupil characteristics which are increasingly recognized as being important to learning. Factors such as pupil self-image, their reactions to expectations, their perception of their teacher's intentions, their status within a group, are recognized as contributing to motivation and perseverance, and therefore deserving of attention when decisions are made about the learning environment. So far, integrated science programmes have not included this kind of information in their evaluation procedures, but there are examples from other fields which show how successfully it can be taken into account. The programme 'Man: A Course of Study' includes a book of 'Evaluation Strategies'[2] for teachers. Five different strategies are proposed for finding out about pupils' understanding, opinions and reactions to the course. Techniques suggested are -

Interview - a way of giving pupils chance to express their feelings.
Classroom Environment Check-lists - for pupils to complete.
Creative Formats - to explore pupils' imaginative and creative work.
Content questionnaire - to diagnose difficulties.
Classroom observing - providing information for change in teaching strategies.

Another project which has provided guidelines for teachers in gathering information about pupil

[1] Progress in Learning Science, op. cit.

[2] Man: A Course of Study, <u>Evaluation Strategies</u>, Cambridge, Mass, Educational Development Centre, 1970.

reactions to the learning situation and in adapting their approach in consequence is the Ford Teaching Project[1]. This project provides a wealth of ideas and evidence relating to the identification and removal of constraints which impede inquiry/discovery learning. The application in science education of these approaches to evaluation and the further development of new ones may well help in devising future integrated science programmes which are able to provide all students with opportunities for intellectual and personal advancement.

Conclusion

Approaches to teaching which are sensitive to differences between individuals aim to provide the conditions which are most conducive to successful learning in all pupils. It has been argued that, when these conditions prevail, then not only are cognitive goals achieved more readily, but the experience of success and the atmosphere of mutual help promote the development of positive attitudes; indeed these approaches provide extremely well for the 'curiosity, compassion and competence', which Albert Baez selected as key words for future integrated science courses.

The essential feature of catering for individuals as interpreted here is that the learning experiences are adapted to the characteristics of the pupils; in approaches where this is not attempted, the pupils have to try to adapt to the learning experiences, and for many this is not possible. Since it is not only the activities but the surrounding organization and the whole setting of learning which affects the pupil, it is necessary to speak of the 'learning environment' as the target for adaptation to individual characteristics.

Research has so far provided few definite guidelines as to the adaptation of various features of the environment to match different aptitudes and modes of learning. Nevertheless, the matching can be carried out in practice by on-going evaluation, the constant gathering of information about pupils and use of this in making decisions about the opportunities and guidance of individual pupils. Thus evaluation has a vital role in providing information as a basis for action and is at the heart of an approach which caters for variations between individuals.

The constraints imposed by local circumstances mean that the approach which seems 'ideal' in one country or one setting may be impracticable in another; therefore approaches have to be devised to operate within these constraints. Some of the child-centred approaches which operate in certain schools in western countries, depending heavily on decisions made by well-trained and perceptive teachers, are not feasible in regions where class sizes are large, teachers have minimum training and resources for a wide choice of activities are limited.

Methods of evaluation which can be used are similarly constrained by circumstances. However it has been shown that, for the purposes of providing immediate feedback, conventional methods of testing are inappropriate; more suitable methods are less structured and capable of use informally, frequently and without subtracting unduly from teaching and learning time. As evaluation is essential to approaches sensitive to individual differences - the approaches so well suited to achieving the objectives which will be increasingly emphasized in the future - it is a matter of urgency to devise or to adapt methods of evaluation to serve this particular purpose.

[1] Ford Teaching Project, Unit 1 Patterns of Teaching, Unit 2 Research Methods, Unit 3 Hypotheses, Unit 4 Teacher Case Studies, Centre for Applied Research in Education, University of East Anglia, 1975.

[2] A.V. Baez, Integrated Science Teaching as part of General Education - Looking Ahead, in Richmond, New Trends ... Vol. II, op. cit.

ANNOTATED BIBLIOGRAPHY

Bruner, J.S., The Process of Education, New York, Vintage, 1960.

Based on the ideas put forward at the 1959 Woods Hole conference in how education in science might be improved in primary and secondary schools, this short book is a discussion of the issues which arise in designing curricula which take into account both the structure of the subject matter and the ways pupils learn. The section in 'Readiness for Learning' deals with kinds of scientific ideas which can be grasped by pupils at different stages of development.

Dean, J., Recording Children's Progress, In the series 'British Primary Schools Today', London, Macmillan, 1971.

A concise outline written for teachers of the purposes of and approaches to keeping records in classes where children are catered for as individuals. Relevant mainly to teachers of children 5-11 years of age.

Engel, B.S., A Handbook on Documentation, Grand Forks, N.D.: University of North Dakota.

A collection of examples of ways in which documentary records can be made about the child, the classroom and the institution. They include records of interviews, anecdotal reports, the use of check-lists, lists of interests and activities, which can be used to build up a bank of information about pupils and their responses to programmes.

Gordon, I.J., Studying the Child in School, New York, Wiley, 1966.

A book which will help teachers who want to find techniques for diagnostic appraisal of their students' and of their own behaviour. Covers non-cognitive as well as cognitive areas, social interactions and teacher behaviour.

Gronlund, N.W., Individualizing Classroom Instruction, New York, Macmillan.

This slim book gives a practical guide for preparing materials which can allow pupils to work at their own pace. Its view of 'individualizing' is somewhat narrow but nonetheless a good starting point. Gives outlines of three approaches: Mastery Learning (Bloom), Individually Prescribed Instruction (IPI) and Programme for Learning in Accordance with Needs (PLAN).

Piaget, J., Science of Education and the Psychology of the Child, London, Longmans, 1970.

Written for non-specialists, this book consists of two major essays by Piaget. The first describes changes on a world-wide scale which have affected education since 1935 and presents a well argued case for giving more attention to the development of the individual. The second discusses principles for education which "appeals to real activity, to spontaneous work based upon personal need and interest" (p. 159). Thought provoking arguments rather than practical guidelines.

Weisgerber, R.A. (ed), Perspectives in Individualized Learning, Itasca, Illinois, Peacock, 1971.

A collection of papers which discuss assumptions which underlie ideas about individualizing learning, some very theoretical, others more practical. The section on 'Individual Differences: How should they be measured and accommodated?' is of immediate relevance to teachers.

10 Evaluation of the African Primary Science Programme (APSP)

Ayotunde Yoloye
University of Ibadan, Nigeria

SUMMARY

This is a description of efforts in curriculum programme evaluation with respect to a primary school science programme (APSP) which operated in over ten African countries. The evaluation programme was as much an experimental programme as the science programme itself.

INTRODUCTION

The African Primary Science Programme (APSP) was one of the three major curriculum development projects in Africa sponsored by Educational Services Incorporated (ESI)[1], Newton, Massachusetts, in the sixties. The other two were the African Mathematics Programme and the African Social Studies Programme.

As a curriculum development project, APSP had a number of unique characteristics.

1. It was a Pan-African programme involving over ten Anglophone African countries. Therefore, the curriculum developers were an international group, and trials of the materials were carried out in several African countries simultaneously.

2. The curriculum developers adopted a truly experimental approach in the development of their materials. Units were written up only after the ideas had been experimentally tried with children in actual classrooms. Such trials took place at international workshops or in science centres in various participating countries.

3. No detailed sets of objectives were formulated before the production of materials. There was only the broad goal of preparing "good" materials for the African classroom.

4. A more detailed specification of goals was eventually formulated with the help of the evaluation team, two years after the programme had been in operation and evaluators had had the opportunity of observing the operation in the trial classrooms.

5. The focus of APSP is the child rather than the discipline of science or the society. Science was visualized simply as a tool for helping the child develop along certain desirable lines.

Overview of evaluation procedure

Concern for evaluation of the curriculum development programme was expressed right from the beginning. At the first international workshop, one member of the team was included to consider problems of evaluation. Subsequent workshops explored evaluation further and more specialists

[1] Later Education Development Center (EDC)

were brought in. The evaluation team that eventually emerged was international as well as inter-disciplinary in nature. Contributors came from the United Kingdom, the United States, Canada and various African countries. They were drawn from education, psychology, sociology, economics and science.

The composition of the evaluation team fluctuated over the years, but the headquarters remained at the University of Ibadan under the direction of the author.

The evaluation procedure was unique in a number of ways. From the beginning, the evaluators realized that because of the emphasis of the curriculum programme on the development of the child rather than the concepts of science, conventional tests of cognitive achievement would be largely inappropriate. There was need to develop novel evaluation instruments and strategies to cope with the innovative feature of the programme. The evaluation procedure therefore became experimental in much the same way as the curriculum development procedure was experimental. A large variety of instruments and strategies of evaluation were tried out, modified or rejected and tried again, and so on. Perhaps one of the most significant outcomes of the evaluation exercise is in fact the philosophy of evaluation that eventually emerged.

Initially, it was believed that the purpose of evaluation was to convince African governments of the worth or otherwise of the programme so that the programme could be adopted or rejected. Within the first two years of the programme's operation, however, it quickly became clear that the essential factors in the adoption or rejection of the programme had little to do with evaluation. The main factors were: (i) the involvement or otherwise of the policy makers from the ministries in the programme; (ii) the absence/presence or quality of competing science programmes initiated or supported by the ministries in the countries; and (iii) diplomacy or the skill of human interaction.

As a result of the above, it was felt that the major role of evaluation should in fact be to inform the curriculum developers themselves of the value of what they were producing and help guide the production so that it could move in the direction of perfection.

This was a significant switch fron <u>consumer oriented</u> evaluation to <u>producer oriented</u> evaluation. A consequence of this switch was that the evaluators thereafter focused mainly on <u>formative evaluation</u>. One implication of this orientation is that the evaluators had to become closely involved in the process of curriculum development. They could not remain simply "objective outsiders". While the assumption of this role to some extent may affect the objectivity of the evaluator, it has the advantage of being less threatening to the curriculum developer and therefore less likely to cripple innovation. Evaluation activities suffered from some major handicaps. The tremendous geographical area involved in the programme made evaluation activities difficult, while there was the difficulty of identifying people with the appropriate expertise and time to work in the evaluation team.

Framework for evaluation

The term "evaluation" is subject to a variety of definitions depending on who is defining it and for what purpose evaluation is being undertaken.

Because of our emphasis on "Formative Evaluation" as indicated earlier, the definition of Marvin C. Alkin was adopted for the purpose of APSP evaluation. He says, "Evaluation is the process of ascertaining the decisions to be made, selecting related information and collecting and analysing information in order to report summary data useful to decision-makers in selecting among alternatives"[1].

[1] Marvin C. Alkin, "Product for improving educational evaluation", <u>Evaluation Comment</u>, September 1970, Vol. 2, No. 3, p. 1.

Table 18. APSP Framework for Curriculum Evaluation.

	Variables		Observation (Instruments and techniques)
ANTECEDENTS (relevant conditions prior to the introduction of the programme)	1. Manpower (teaching, supervisory developer etc.)	- Quality, quantity, qualifications, predispositions	1. Questionnaire surveys, direct observation, tests
	2. Children	- Intellectual ability, socio-economic background, cognitive styles	2. Questionnaire surveys, tests
	3. Learning environment	- School setting, educational administration economic resources, political and cultural setting	3. Direct observations, questionnaire surveys
TRANSACTION (Various kinds of interactions and activities involved in the development and presentation of the programme)	1. Instructional materials	- Nature, quality, cost and development strategies	1. Readability indices, task analysis, cost effective analysis, expert judgement
	2. Instructional techniques	- Teacher-child-material interaction	2. Direct observation using observation instrument e.g. interaction analysis
	3. Supportive strategies	- Teacher training, interaction with policy makers, interaction with resource people	3. Direct observation.
	4. Intervention strategies	- Sequence of introducing materials into schools	4. Survey
OUTCOMES (the effect of introducing the programme on various categories of people, organizations, policies, etc.)	1. Cognitive changes in	Child, teacher, parents, curriculum developers, administrators	1, 2, 3 and 4. Test, direct observation, designed experimental studies, surveys
	2. Affective changes in	Child, teacher, parents, curriculum developers, administrators	
	3. Psychomotor changes in	Child, teacher, parents, curriculum developers, administrators	
	4. Long-term effects on life of children, e.g. occupational choice, earning capacity, adjustment to society		
	5. Effect on educational administration and policy		5. Surveys, direct observation
	6. Economic implications		6. Surveys, designed experimental studies, cost-effective analysis

A framework for evaluation based on a model proposed by Stake[1] was drawn up to define the full scope of evaluation and identify procedures for collecting evaluation data.

Table 18 shows the framework, with a list of various evaluation instruments and techniques employed in data collection. In many cases, these had to be designed specifically to fit the APSP situation.

The task of evaluation, based on our definition, was threefold:

1. Helping to identify and describe in appropriate language, the curriculum developer's assumptions with respect to antecedents and intents with respect to transactions and outcomes.

2. Designing and using a variety of techniques and instruments to collect data with respect to the validity of these assumptions and the realization of the intents. The task of evaluation in this latter respect, however, went beyond merely checking whether intents were realized or not, to identifying unintended outcomes.

3. Analysing the data collected and supplying summary information to the curriculum developers, organizers, etc., to help in the process of decision-making.

Identification of goals

It is common to expect that curriculum developers should themselves carry out the task of describing their assumptions and intents, so providing a clear list of objectives which the evaluator can use. Experience shows, however, that the curriculum developer is often not equipped to do this. The evaluator therefore has to work in close co-operation with the curriculum developer in order to evolve a usable description.

In the case of APSP, the evaluation group adopted an indirect technique of getting the curriculum developers to clarify their objectives. On the basis of direct observation and analysis of the programme's materials and activities (the programme had by this time operated for two years), a list of objectives at which the programme seemed to be aiming was prepared. This list was compiled into a rating scale which was then circulated among participants at the workshop. Each participant was required to rate the degree of agreement with each stated objective on a five-point Likert-type scale. Space was provided for explanatory comments and additional suggestions. On the basis of the responses to this rating scale, a summarized list was prepared of the broad goals with which all participants seemed to agree. These goals were in two parts. The first dealt with intended outcomes and the second with intended transactions.

Part I

APSP aims at developing the following characteristics in children:

1. first-hand familiarity with a variety of biological, physical and man-made phenomena in the world around them;

2. interest in using initiative for further exploration of the world around them;

3. ability to find out for themselves - to see problems and to be able to set about resolving them for themselves;

4. confidence in their own ability to find out for themselves and to do things for themselves;

5. ability to share in a common development of knowledge through collaborating on problems, telling, listening and discriminating use of secondhand sources.

[1] Stake, "The countenance ...", op. cit., p. 538.

Part II

In order to foster the above development in children, APSP science should have the following characteristics:

1. the focus of study should be on concrete phenomena themselves;
2. the materials selected should capture and hold the attention and interest of children;
3. the materials should reveal that there is not always one right answer;
4. materials should allow opportunities for a variety of different ways to find out;
5. the classroom experience should lead to social interaction among children;
6. to a large extent the materials should be simple and familiar;
7. the materials should encourage children to do things on their own, in their own ways.

This is as detailed as APSP ever went in defining its objectives. Did this worry the evaluators? Not really. The choice of the child for emphasis implies attention to such powerful psychological phenomena as "individual differences", "readiness" and "learning climate". It consequently implies a predominantly "open" classroom where the child's participation in deciding what is to be learnt is high, and in which, therefore, a catalogue of what will be ultimately learnt (especially in the cognitive domain) is impossible.

Evaluation of APSP philosophy

Given the theoretical and philosophical orientation of APSP, are there any objective ways of telling if this orientation is a valid one?

The evaluators examined this question and provided for the curriculum developers a summary of empirical data that validate the orientation as well as grounds for doubting the validity.

As with any curriculum emphasizing the child factor, empirical evidence of validity is to be found in psychology. The work of Piaget provides some of this evidence. The fact that the majority of primary school children will be at the concrete operational stage validates APSP's insistence that the focus of study should be on concrete materials and phenomena. The finding that passage from one stage of intellectual development to a higher one may be hastened by appropriate experiences with a variety of materials, validates the APSP practice of encouraging free play with a variety of materials like water and sand, in the earliest classes. The finding that a child's basis of interpreting natural phenomena is usually very different from that of adults (and alters with age) validates the APSP insistence that children be allowed to do things in their own way. The work of Maslow, Ward and others on the concept of "learning how to learn" validates APSP's insistence that materials should present a variety of problem situations, with the possibility of a variety of answers and variety of ways to find out. The work of Anderson, Lewin et al, and Flanders, on classroom climate validates APSP's insistence that teachers be predominantly integrative rather than dominative.

Three grounds for doubting validity were identified. First, Beeby's[1] hypothesis of educational stages typifies one kind of basis for doubt. He hypothesized four stages of growth of a school system, namely, (i) dame-school, (ii) formalism, (iii) transition, and (iv) meaning. Each stage is characterized by certain characteristics of teacher and curriculum. Characteristics of teacher range from ill-educated, untrained in the dame-school to well-trained in the meaning school. Curriculum characteristics vary from unorganized, relatively meaningless symbols

[1] C.E. Beeby, The quality of education in developing countries, Cambridge, Mass., Harvard University Press, 1966.

requiring rote meanization in the dame-school, to the emphasis on meaning, understanding, activity and problem-solving methods in the meaning school. APSP curriculum is striving in science for stage (iv) - meaning. The basis for doubt is whether this goal is realistic given the present situation in Africa where the majority of teachers in the primary school are in fact at stage (i) or (ii) - dame-school and formalism.

Next, many African countries' policy-makers in education in fact emphasize society in their pronouncements, on curriculum. The school is seen as an agent of economic development and modernization. Consequently, educational policies are guided by political and economic considerations. The basis for doubt is whether a child-oriented curriculum like APSP will be perceived as relevant to a society-oriented educational philosophy.

A third basis for doubt directly challenges the psychological validity which was earlier established for APSP. This is a query as to whether psychological findings based mainly on research in western societies are valid for African countries and whether socio-cultural factors in African countries would not confound the validity of such findings.

Implications of evaluation of philosophy

The last two points need to have research carried out into them. There already exists a number of studies by African and other scholars on the cognitive development of African children (Etuk, Otaala, Okonji, Ohuche, Duruji, Jahoda) which show that basically the findings of Piaget hold true in African children.

It is however the first point that has provoked major action on the part of APSP (now SEPA). In terms of decision-making APSP had two alternatives open. One was to modify the curriculum orientation to suit the characteristics of teachers. The other was to change the characteristics of teachers to fit the new orientation. APSP chose the latter alternative and since 1971 its major thrust through SEPA has been in teacher education. A trial handbook for teachers has been produced, a teacher's resource book is in the making and, perhaps most significant, a centre for experimenting in the education of teachers for primary science teaching has been established in University College, Njala. Trainees are drawn from all over Africa.

Evaluation of antecedents

It was not necessary to conduct fresh studies into every one of the antecedent variables. A combination of three approaches was used. First, intuitive evaluation: the strategy of APSP curriculum development, whereby no unit gets written up until the ideas have been thoroughly tried in actual classrooms, was such that much of the evaluation of antecedents (e.g. administrative set-up, economic resources, school setting, socio-cultural setting, and children's and teachers' predispositions was done intuitively on the job by the science educators and their assessments influenced their choice of elements to include in the units. Second, existing publications including official publications such as annual statistics of education, annual reports of education, development plans, reports of special commissions of education, and political policy statements, were studied to provide relevant information on antecedent conditions. Other existing publications were specific studies, such as surveys of socio-economic background[1] and a survey of conditions in

[1] P.J. Foster, Education and Social Change in Ghana, London, Routledge and Kegan Paul, 1965; and E.A. Yoloye, "Socio-economic background and school population. A survey of the background of children in three types of schools in the western state of Nigeria", The Education in New Countries, vol. 12, no. 1, 1971, p. 5-18.

schools [1]. Finally, there were some specific studies carried out in connexion with APSP [2].

Some of the findings that are directly relevant to APSP include the following: general intellectual ability and manual dexterity are the most critical factors in determining favourable teacher response to the use of the APSP approach; previous knowledge of science is not absolutely essential for teachers to use the APSP approach; age and teaching experience correlated positively with favourable teacher response to the use of the APSP approach; the reading abilities of the average primary school teacher are high enough for them to comprehend the majority of APSP unit teachers' guides; and the reading abilities of the average primary school child are such that the children may have difficulties with their background readers.

Evaluation of transactions

Because of the emphasis on formative evaluation, the bulk of evaluation activities was concentrated on developing and trying out instruments and techniques for transactions.

Evaluation of instructional materials

Because of the open-ended nature of APSP approach, it is not always easy to identify precisely what learning experiences the units are supposed to be providing. One task of evaluation was therefore helping to identify what each unit is likely to do. Two approaches were used in determining the potential of the various units.

Task analysis: this process, a modification of Gagne's technique, sets out to identify as precisely as possible the concepts which a unit sets out to teach and the experimental operations, observations, and logical operations involved in the process of teaching them. It differs from Gagne's procedure in that it is carried out on the written units rather than as a foundation for preparing the units. Strictly speaking, the technique is more suitable for structured, concept-oriented curricula than for the open-ended approach of APSP. However, APSP units vary in their degree of structure and concept orientation. Some units are sufficiently structured and concept-oriented to be amenable to task analysis, but most APSP units would not lend themselves to such analysis.

When applicable, task analysis serves two main purposes: it helps to identify flaws in sequencing of various experimental operations and spotting omissions or inappropriate processes and thus helps in the writing exercises of a unit and it provides a basis for the setting of achievement tests.

Expert judgement on the potential of APSP units to teach the processes of science: fifteen scientific processes - a modified form of the American Association for the Advancement of Science (AAAS) Science-A Process Approach list - were selected. Each APSP unit was then rated by a panel of eight to ten "experts" in science on a five-point Likert-type scale on its potential for teaching each of the fifteen processes. Ratings were then pooled. Whenever there was wide divergence in ratings, there was a thorough discussion in which each rater showed reasons for his or her rating. Such discussions usually ended in closer agreement of ratings. In addition, each

1 D. Calcott, "Education in a rural area of western Nigeria", (Mimeo) I.L.O. Interim report, Ministry of Education, Ibadan, 1968.

2 E.A. Yoloye, "A study of teacher response to training on APSP materials", Journal of the Science Teachers Association of Nigeria, vol. 9, no. 1, 1970, p. 21-25; E.O. Ogunyemi and F.M. Eboda, "Cognitive preferences among high and low physics achievers in two Nigerian secondary schools" and E.O. Ogunyemi and G.L. Bettie, "Cognitive preferences in mathematics among high and low mathematics achievers in two Nigerian secondary schools", African Journal of Educational Research, vol. 1, no. 1, 1974, p. 107-114 and 97-106; and E.A. Yoloye, "Readability indices in the evaluation of curriculum materials", Journal of Curriculum Studies, vol. 7, no. 1, 1975, p. 78-84.

unit was rated as to its suitability for upper, middle or lower primary classes. Profiles were then constructed for each unit.

On the basis of these profiles, a number of major observations emerged. All units analysed appeared appropriate in terms of difficulty and complexity for primary school children. The profiles of each unit could form a basis for selection of what units to combine for a curriculum so that the entire curriculum is balanced in terms of the processes catered for. Some processes do not seem to be catered for sufficiently by the nineteen units analysed. In particular, there is little or no indication of making operational definitions and formulating models. Others which are only scantily catered for include prediction, inference, formulating hypotheses and interpreting data.

Readability of APSP materials

Another kind of evaluation carried out on the APSP materials was directed at assessing how easy it could be for the audience (teachers and children) to comprehend the materials written for them.

The Lorge formula for estimating the difficulty of reading materials was used for calculating the readability indices of thirty-seven APSP units. For comparison, the readability of ten Elementary Science Study (ESS) units and three Nuffield junior science materials were also calculated.

Taken along with the data on reading ability of primary school teachers and children, the data on readability indices provided valuable insight into the probability of the materials being appropriate for the teachers and children.

It was found that the teachers' guides were all within the range of comprehension of the majority of Grade II teachers. The most difficult unit has a readability index of 6.84 while the twenty-fifth percentile for Grade II teachers is 7.15. In addition, for the majority of primary schools, there may be difficulty with the pupil materials. Five of the eight books analysed had readability indices of 4.83 or higher. At least half of primary 6 children in the sample tested (median reading level 4.73) would have difficulty with these materials. The mean readability indices of ESS and Nuffield materials analysed are 6.07 and 6.59 respectively as compared with 5.35 for APSP materials. This indicates that the APSP writers were in fact sensitive to the reading abilities of their audience.

Evaluation of instructional strategies

Part II of the goals of APSP refers to transactions. They suggest certain instructional strategies which may be summarized as follows:

- children should be learning more from concrete materials than from the teacher. Therefore, for most of the time the science class should consist of activity with concrete materials by the children;
- much of this activity should be on individual or small group basis;
- the teacher should be playing a supportive, integrative role rather than a dominant role; and
- the interaction of the teacher with the class at most times should be through individual children and small groups of children.

The task of evaluation with respect to teaching strategies was two-fold. First, to help identify appropriate patterns of class organization and teacher behaviour for teacher-training purposes and second, to provide a basis for assessing correct interpretation of the APSP approach for the purpose of evaluating the programme.

Three instruments were developed for these purposes:

<u>The class interaction</u>: this is used for recording objectively the pattern of interaction of the teacher with the whole class. It is a category system of direct observation making use of five categories:

A. group or individual activity;
B. whole class discussion;
C. transition;
D. teacher lectures;
E. other.

Records are made at one minute intervals although smaller intervals (e.g. thirty or fifteen seconds) have also been used. At the end of the lesson, an interaction profile is drawn based on the number of entries in each category.

If APSP approach is being correctly interpreted, a profile showing the highest point on A followed by B but with very low values on C, D and E is expected. For teacher-training purposes, records of a teacher's performance either in micro-teaching situations or normal class situations can be used to monitor the teacher's progress towards the expected pattern. The profiles, coupled with sequential pattern of the entries, also provide fruitful bases for initiating discussion with the teacher.

For programme evaluation purposes, the instrument is useful for identifying "experimental" teachers whose classes will be compared with control ones.

<u>The verbal interaction sheet</u>: this is designed to record objectively the teacher's pattern of verbal interaction with the class. Much of the stimulation of the child is through the teacher's verbal statements. Such verbal statements could foster or hinder the creative thinking and activity of the child. The verbal interaction sheet identifies seven categories of verbal statements, namely:

A. direct instruction;
B. open-ended question;
C. close-ended question;
D. suggestion;
E. giving information;
F. supportive action;
G. comment implying that the child's statement is unacceptable.

Entries are made in the sequence in which they occur and a profile is constructed at the end of the lesson.

The pedagogic objectives of APSP place premium on statements in categories B, D and F. Therefore, an ideal profile should show peaks at B, D and F and valley at A, C, E and G.

As in the case of the class interaction sheet, this instrument can be used for teacher-training purposes, in micro-teaching or normal class situations, and for identifying teachers with appropriate verbal interactions for experimental classes in programme evaluation.

<u>Waiting time sheet</u>: APSP emphasizes the need to let the child work at his or her own pace. During question-and-answer exchanges between teacher and child, the teacher is often impatient to have the right answer from the child, with the result that slow-thinking children in particular, and indeed children in general, become inhibited in their thinking processes as a result of the teacher interrupting with another question or statement before they have had time to think through one question.

The waiting time sheet provides an objective record of how long a teacher waits after each question before speaking again. The records also provide some evidence of the extent to which children are inhibited by teacher talk.

This instrument is specifically for teacher-training purposes. It is particularly effective in micro-teaching situations for monitoring the progress of the teacher towards greater patience in handling children.

Evaluation of supportive strategies

Supportive strategies include in-service courses or workshops, follow-up of teachers in schools and interaction with policy-makers and resource people.

Evaluation in these respects focused on developing effective feedback systems so that supportive strategies may be progressively improved. APSP used a variety of questionnaires for this purpose including: reactions to a training course; report on teaching of science units; and materials information sheet.

Evaluation of outcomes

Two summative evaluation studies were carried out with respect to some outcomes of the programme on children and teachers.

Outcome on children: Duckworth[1] compared children in fifteen primary school classes which had been exposed to the APSP materials for periods varying from two terms to two years, with children in thirteen primary school classes which had not been exposed to APSP in Kenya on a variety of tasks of intellectual ability and creativity. The tests of intellectual ability included: missing piece, straight line, bilo, lego corner, and ordering weights.

For creativity, a set of materials equally familiar or unfamiliar to the two groups of children was presented to the children, and they were asked to do whatever they liked with them. They were then rated by observers in terms of the number of different things done and the complexity of things done.

In general, APSP children did significantly better than non-APSP children on all tests.

Outcome on teachers: Johnson[2] compared six APSP classes with six non-APSP classes in Sierra Leone on an interaction analysis schedule based on the Flanders but having twenty behaviour categories. Five consecutive lessons were observed in each class, with findings as follows:

- in APSP classes, the incidence of pupil activity was higher than teacher activity whereas the reverse was the case in non-APSP classes;
- APSP teachers used predominantly indirect influence whereas non-APSP teachers used predominantly direct influence;
- pupils in APSP classes exhibited more independent behaviour than non-APSP classes; and
- length of contact with APSP did not seem directly related to the amount of change in teacher behaviour.

Concluding remarks

The evaluation experiences of APSP raises crucial issues of criteria of success of a curriculum development programme. Should it be the quality of the materials, the degree of adoption of the materials or the quality of people produced by the programme?

[1] Eleanor R. Duckworth, A comparison study for evaluating primary school science in Africa, Newton, Mass., E.D.C., 1971.

[2] Victor L. Johnson, "An investigation into the effect upon the classroom interaction of the introduction to Sierra Leone of the African primary science programme", (Mimeo) Newton, Mass. E.D.C., 1970.

In the course of APSP evaluation, it became clear that the crucial criterion for assessing the success or otherwise of APSP cannot be merely "whether or not the materials work in the African classroom" but more importantly "whether or not APSP has contributed significantly to the production of a core of people within Africa, with experience, expertise and a mental orientation geared towards continuous curriculum renewal".

There is now a substantial core of such people all over Africa. The harnessing of the resources and expertise of this core of people for joint effort at tackling curriculum problems in various countries has been carried out in a most effective manner by the Science Education Programme for Africa (SEPA).

Not only has this organization promoted continuous training of manpower through two international institutions - The International Centre for Educational Evaluation (ICEE) at the University of Ibadan, Nigeria, and the Centre for Science Education in Njala University College, Sierra Leone; it has also made it possible on several occasions to assemble the best available human resources in Africa to work together on in-country programmes in curriculum development and evaluation.

In fact, SEPA may be said to have evolved a unique model of educational development in African countries through its emphasis on manpower development and harnessing of expertise with a minimum of administrative staff. Since its inception in 1971, SEPA has organized over a dozen international workshops, training programmes and seminars all over Anglophone as well as Francophone Africa.

ICEE has trained more than thirty people in educational evaluation in the past three years and has a current intake of ten for a one-year diploma. It has also conducted four major curriculum evaluation studies at national and international levels. The Science Education Centre in Njala has produced its first set of one-year diploma students in science education.

In terms of the materials produced by APSP, they have not been adopted wholesale in any country. However, they have formed substantial inputs into national curricula in most of the participating African countries including Nigeria, Ghana, Liberia, Sierra Leone, Kenya, Uganda and Tanzania. The fact that no wholesale adoption has occurred is probably a positive thing because, ironically, if the programme has produced the right orientation among its exponents, one test of its success would be that in the coming years, its materials should be virtually completely replaced by new materials produced by the very people who have had contact with APSP.

ADDITIONAL READING

Robins, Ralph H., Programming for Change in Science Education, APSP Monograph Series, Newton, Mass., E.D.C., 1970.

Yoloye, E.A., Evaluation for Innovation, Ibadan, I.U.P., 1976.

11 Summative evaluation of Caribbean integrated science projects

Judith F. Reay
University of the West Indies, Trinidad

SUMMARY

In the Caribbean in the early seventies, there was a complex situation in which several integrated science innovations were taking place in the region. The major strategy for summative evaluation was based on an educational experiment and upon achievement tests, but other forms of evaluation have also been used. There is little evidence that the data were used by the decision-makers, and it is suggested that, in the end, decisions are taken on the basis of availability of published materials and individual appraisal of these. Teacher education courses should pay more attention to the significance of evaluation data.

INTRODUCTION

This is an account of the combined evaluation of two concurrent curriculum development projects, the MONA Project of the University of the West Indies (UWI) in Jamaica and the CEDO/UWI Caribbean Regional Science Project based at the University of the West Indies in Barbados. To help clarify the issues raised by this exercise, we shall also look briefly at other attempts at summative evaluation of these curricula and of the original WISCIP of the University of the West Indies in Trinidad.

In planning for the evaluation of a curriculum development project, one has to take decisions on the following questions:

- what is the purpose of the evaluation?
- what can (and should) be evaluated?
- what procedures will be used?

As Cohen points out in the introductory chapter of this book, the answers to these questions will be heavily dependent on the context of the evaluation. This case study describes an evaluation in a setting where the sort of neat curriculum development strategies outlined by Tyler[1] and Kerr[2] would not succeed. Time was short, resources were small and sensitive issues were numerous. It was thought that the most powerful evidence for the decision-makers (for whose guidance the evaluation was intended) would be children's achievement in tests. Since evaluation is for decision-making (Welch, chapter 3), the summative evaluation had to be of a nature which would reflect these considerations.

We shall discuss later whether the evaluation exercises did in fact lead to sound decision-making.

[1] Tyler, op. cit.

[2] J.F. Kerr, "Curriculum change in emergent countries", in G. Howson (ed), Developing a New Curriculum, London, Heinemann, 1970.

The purpose of the evaluation

In the Caribbean in the early seventies, there was a rather complicated situation in science curriculum development for the lower secondary level. UWI at Trinidad was completing the development of WISCIP; UWI at Jamaica was continuing the development of a programme that had been begun by the National Curriculum Development Committee (the UWI phase being known as the MONA Project); and UWI at Barbados was the headquarters of the CEDO/UWI Caribbean Regional Science Project embracing fourteen countries of the region including Jamaica and Guyana (but not Trinidad and Tobago). Why were there three such projects, all co-ordinated by UWI?

The explanation is that at one and the same time, the West Indies is a region and it is fifteen countries with fifteen Ministries of Education, a situation that must be familiar to the South Pacific and to Africa. The separate governments of the region established junior secondary schools within a few years of one another so that secondary education would be available to more children than had been served by the selective grammar schools; it is no coincidence that the curriculum development projects sprang up at about the same time.

The junior secondary schools are generally subject to considerable government direction, although some of their teachers are very ready to criticize the curriculum and suggest improvements. The grammar schools are much more autonomous, but many of them choose to use the same curricula as the junior secondaries, for good reasons.

Thus the governments and their junior secondary schools were very interested in the science curriculum developments, as were many of the grammar schools. Other schools, reasonably satisfied with their own programmes and reluctant to offer their children as guinea pigs, were cautious about participating in any project.

An obvious purpose of the evaluation was the accountability to funding agencies and governments. However, the actual form of the major evaluation was probably determined by Jamaica, where both MONA and the Regional curriculum were in trial. Thus an important purpose there was to help teachers, the government and the curriculum committee make a choice between the two curricula.

What outcomes would be evaluated?

Curriculum innovation has a variety of possible effects. Teachers can become more knowledgeable about the subject they are teaching and they can become more skilled at helping children to learn. Children can acquire skill in the processes of science, they can grasp the major concepts of science better, and they can develop attitudes such as self-esteem, understanding of and concern for other people and a love of their environment. Children and teachers can become excited by the teaching-learning situation. Parents and other members of the community can develop a better perception of what science learning is all about and can be more willing to co-operate with children in their out-of-school investigations.

Some of these outcomes are exceptionally difficult to evaluate, but all of them were goals of the projects. However, teachers in the region at that time were more concerned with cognitive achievement than with anything else. A programme which was superior in everything except the examination results which it produced would be a rejected programme. Resources were small and time was short, so the project workers decided, rightly or wrongly, that their major summative evaluation would attempt to compare the development of cognitive skills. Tawney (Chapter 2) has stressed that the evaluator is the servant of the decision-maker; in the Caribbean exercise, it seemed important (in view of the complicated and sensitive issues) to carry out an evaluation that was as objective as possible and that would provide the information sought by the decision-makers.

The procedures used

Setting up an experiment

The Caribbean Regional Science Project (CRSP) was important, serving many of the schools in Barbados, ten schools in Jamaica, some of the schools in Guyana and Belize and nearly all the

secondary schools in the smaller islands. The MONA[1] Project, which began at about the same time (in late 1970), was much smaller, being only part of a science teaching improvement project. It was decided that Jamaica should set up a CRSP/MONA control experiment. Matched against the Jamaican CRSP trial schools would be a set of MONA schools as similar as individual organisms can ever be. Because of the limited resources of the Jamaican project, the third set of schools (the controls) would not receive special attention until the time came for the summative evaluation. This was not considered a serious weakness in the experimental design. The improvement project had, after all, many activities and their influence on all schools would be taken into account when selecting the controls.

Communications with trial schools can be a problem; in Kenya's SSP and Guyana's SDSP, for example, trials have been restricted to schools close to the curriculum development centres. In Jamaica, for experimental as well as for educational reasons, it was thought important that the trial schools should cover a broad spectrum - rural and urban, fishing village and banana-growing areas, selective and non-selective. The project's resources were stretched rather thin by having trial schools in most corners of the country. The classification of the trial schools is shown in Table 19.

Table 19. Classification of trial schools in Jamaican curriculum experiment.

		MONA	CRSP
Selective		2	3
Non-selective	- Comprehensive (poor facilities)	-	2
	- Junior secondaries		
	. World Bank (good facilities)*	5	5
	. Converted (poor facilities)	1	-
Total		8	10

* Of these, three tried out both curricula in different streams.

The schools in both samples were matched as carefully as is possible in an educational experiment, which, being concerned with people, must lack the rigour of a non-human laboratory experiment. Many of the variables could be taken into account when the summative evaluation was carried out.

Preparing for summative evaluation

The Caribbean projects have all had short development stages - about three years. While this leads to a certain untidiness in procedures when a three-year curriculum has to be designed, subjected to formative evaluation, modified several times and then finally published, disseminated and implemented, there is something to be said for generating curricula quickly. Commonly, curriculum development projects have been set up in response to an urgent need and the nation is anxious to get its hands on the published materials. At least in the Caribbean, teacher mobility is very high, and all kinds of problems arise in the trial schools when teachers come and go. Teachers get tired of sending in regular feedback. It is common for funding of a project to be promised for only three years at a time and the decision to renew, if it happens, always seems to be taken at the eleventh hour.

[1] J.F. Reay and A.D. Turner, New World Science. (The published version of the MONA curriculum.) Pupils' books and teachers' guides, Trinidad, Longman Caribbean, 1975 and 1976.

With such short projects, it is necessary to conduct summative evaluation during the development stage. Data were collected about the performance of children after they had been following the courses for one and two years, but it was not possible to examine performance after a full three-year course.

There was also the question of who should carry out the evaluation. Williams[1] has commented that evaluation requires a certain detachment from the day-to-day detail of curricula. Tawney observed that, as far as summative evaluation is concerned, self-report has low credibility. The projects lacked the means to employ any evaluators; however, independent experts in the region were approached. It was hoped that one of them would take responsibility for analysing and interpreting the data. Perhaps because of the delicacy of the whole situation, this kind of help was refused, although the experts were very generous with advice. Thus the innovators were also the evaluators, which had the advantage of costing nothing. In the end, however, this may have reduced the influence of the exercise on decision-making.

The instruments of evaluation presented a difficulty. Mayer (chapter 6) stresses that the quality of the information gained is only as good as the quality of the instrument, and argues the need for better co-ordination in the development, repeated use and refinement of instruments. In the Caribbean, searches failed to turn up any instruments which would be suitable for West Indian children and which would measure specific cognitive skills. As it later turned out, Williams[2] had been developing achievement tests for WISCIP in Trinidad and Tobago and developing analytical procedures for the responses. The potential of his work for testing a variety of skills on a large scale and with small resources is very significant. Had the results of his work been available at the time of evaluating the experiment, they would probably have influenced the tests developed for the eastern and northern Caribbean. As it was, original instruments had to be constructed. Pre-testing them (a difficult and expensive process, as Mayer says) was beyond the resources of the projects; but some post-test inspection was done and the validity and reliability were thought acceptable[3].

The samples

Although it was the situation in Jamaica that had led to the establishment of an educational experiment, the summative evaluation embraced eight countries in seven of which the only experimental curriculum was that of CRSP. (CRSP's curriculum should be clarified here. A descendant of Trinidad's WISCIP, its first version was called WISCIP/B. Formative evaluation all over the Caribbean led to the development, in regional writing workshops, of WISCIP/C. This version was further modified to become the curriculum published as WISC, West Indies Science

[1] I.W. Williams, WISCIP: A Case Study in Curriculum Innovation, London, Commonwealth Secretariat, 1973.

[2] I.W. Williams, Report on the Year 1 WISCIP Examination set in Trinidad, Summer 1970. Swansea, University College, 1971 (Unpublished); and Report to Institute of Education, University of the West Indies, on the 1971 WISCIP Examinations (Years 1 and 2), held in Trinidad in July 1971, Swansea, University College, 1971 (Unpublished).

[3] J.F. Reay and A.D. Turner, An Evaluation of the First Year of Two Science Pilot Projects on Trial in Jamaica, Jamaica, UWI, 1972 (mimeo); P.S. Adey, J.F. Reay and A.D. Turner, An Evaluation of New Junior Secondary Science Curricula in the Caribbean - Interim Report, Jamaica, UWI, 1973 (mimeo), and The Caribbean Science Curriculum: Evaluation Test 1972, Barbados, UWI, 1973 (mimeo); and P.S. Adey, L. St. E. Jones and A.D. Turner, An Evaluation of New Junior Secondary Science Curricula in the Caribbean - Second Report, Jamaica, UWI, 1973 (mimeo).

Curriculum[1]. The MONA curriculum has been published as New World Science, 1975.)

From the experimental groups, equivalent samples were selected. The selection of controls presented no problem in Jamaica, where most of the schools were not participating in these innovations, although some were using other new curricula (e.g. Nuffield). In Jamaica, therefore, a respectable battery of controls faced the project schools. Barbados had rather more difficulty in selecting control schools and no suitable ones could be found in the other six countries.

The costs

When the results were processed by computer (1972 and 1973), the charge for the UWI computer service was of the order of (US)$100 each time. Materials costs were reduced by re-using the test papers, but in any case such expenses as paper and travelling, which are part of the normal innovation process, were not charged separately to evaluation. Hence, the only costs specific to the evaluation process were those of the computer.

The tests and their administration

During the lives of the projects (1970-1973), three waves of testing took place.

<u>1971. MONA and WISCIP/B in Jamaica</u>. An objective test was constructed with small sections to test a variety of skills. From precise instructions, teachers had the papers typed and marked the responses. The only change they could make to the items was to select a stated number from the recall sections to suit their own courses. The whole procedure lacked some rigour in selection, presentation to children, and marking, but analysis revealed no serious weaknesses.

The test was administered to 2,283 Form 1 children of whom 710 were MONA pupils, 611 were WISCIP/B pupils and 962 were controls.

We shall compare the performances of these three groups later (Table 21), but there were other interesting findings[2]. As with all subsequent tests, children in selective schools performed better than those in non-selective (comprehensive and junior secondary) schools. Had it been otherwise, the instrument would have proved invalid. There was no difference in performance in the Cheap Recall section, whether comparing children of different abilities or comparing children following different courses. This evidence has proved to be of great interest to teachers who, while accepting the worth of a new curriculum in principle, retain doubts about their pupils' preparation for an external examination, which they perceive as a test of facts. This was perhaps the most convincing and useful information obtained from any stage of the evaluation exercise, and it is easily replicated by any teacher.

[1] P.S. Adey (ed), West Indies Science Curriculum, London, Heinemann, 1976.

[2] Reay and Turner, An evaluation ...., op. cit.

1972. MONA and WISCIP/B in eight countries. The same wave of children, now Form 2, was examined. Experience with the 1971 testing had suggested that tighter procedures were needed.

A draft test paper, prepared by the headquarters teams, was shredded, modified and reclassified by independent and highly qualified experts. The categories of cognitive skills to be tested separately were Comprehension, Application and Higher Abilities, and the headquarters teams were confident that knowledge components had been minimized. Once more, parallel groups of children were tested, as shown in Table 20.

Table 20. Samples for 1972 summative evaluation by Jamaican science project and Caribbean regional project.

	MONA	WISCIP/B	CONTROL	TOTAL
Jamaica	371	470	485	1,326
Barbados	-	364	375	739
Six other countries	-	741	-	741
Total	371	1,575	860	2,806

The test was administered by the curriculum workers themselves, according to carefully agreed procedures, and children responded on mark-sense computer cards. The human variability in administration and marking was thus virtually eliminated. Performances are compared later (in Table 21).

1973. WISCIP/B and WISCIP/C in eight countries. It was now time to evaluate CRSP's modification of WISCIP/B to WISCIP/C. The same instrument as in 1972 was applied in the same way to 809 of the second generation of Form 2 children. On this occasion, it was administered to only WISCIP/C pupils, selected in such a way that data could be compared with WISCIP/B and control results of the previous year.

The information obtained

After the first year of the innovations in Jamaica, Reay and Turner[1] showed that non-selective MONA children performed better than the controls, selective MONA children gave the same performance as the controls, and WISCIP/B children performed worse than the controls.

However, with the same wave of children a year later, Adey, Reay and Turner[2] showed, inter alia, that in Jamaica and Barbados all groups of project children performed better than the controls, except for some non-selective children (MONA in Jamaica and WISCIP/B in Barbados) whose performance was the same as that of their controls.

When the next wave of children (those who had followed WISCIP/C) were tested, their performance was better than or equal to performance with WISCIP/B, except for non-selective children in Jamaica and St. Lucia, whose performance was worse[3].

From these three studies, it has been possible to compile Tables 21 and 22, which summarize some of the results.

[1] Reay and Turner, An evaluation ..., op. cit.

[2] Adey, Reay and Turner, An evaluation ..., op. cit.

[3] Adey, Jones and Turner, op. cit.

What escaped the workers at the time is the lack of consistency revealed by this longitudinal study. Taken one year at a time, "hard facts" were collected at respectable levels of significance. But when one puts them all together and looks back with the increased objectivity that comes with a time lapse, it is difficult to maintain that anything has been proved. The tables certainly show more pluses than minuses, but there is a lack of consistency from one year to the next that casts serious doubts on the validity of the results.

Table 21. Comparison of mean scores on achievement tests: trial children against controls.

		1971	1972	1973
MONA (Jamaica)	- selective	=	+	
	- non-selective	+	=	
WISCIP/B	- selective (Jamaica)	-	+	
	- selective (Barbados)		+	
	- non-selective (Jamaica)	-	+	
	- non-selective (Barbados)		=	
WISCIP/C	- selective (Jamaica)			+
	- selective (Barbados)			+
	- non-selective (Jamaica)			=
	- non-selective (Barbados)			+

\+ better than controls
\- worse than controls
= no significant difference from controls

Table 22. Comparison of mean scores on achievement test: WISCIP/C (1973) children against WISCIP/B children (1972).

	Selective	Non-selective
Antigua	=	=
Barbados	=	+
Dominica	+	-
Grenada	=	no results
Jamaica	=	-
St. Kitts/Nevis	none	+
St. Lucia	+	-
St. Vincent	+	+

\+ WISCIP/C better than WISCIP/B
\- WISCIP/C worse than WISCIP/B
= no significant difference between WISCIP/B and WISCIP/C

One might ask if cognitive development can be measured by an achievement test lasting for one or two hours. One might go further and reflect that such events as teachers' maternity leave (and most Jamaican teachers are women), or the government's failure to complete its new laboratories to schedule, or even an important series of cricket matches, might be more influential than the curriculum.

Other summative evaluation procedures

Tawney (chapter 2) suggests that triangulation (the use of several methods) can improve the reliability of the information obtained, and Cohen (chapter 1) has outlined several possible methods.

Some postgraduate students at UWI in Trinidad have been using the test developed by Adey, Reay and Turner[1] in small studies which compare the performances of Trinidad WISCIP children, Trinidad control children and other groups of children in the Caribbean; but no important information has emerged.

Williams[2] has been analysing test scores of Trinidad WISCIP children and has found that some curriculum objectives are apparently not being achieved. He casts doubt on the objectives themselves (Tawney, in chapter 2 of this book, has pointed out that one must think seriously about this) and observes that a number of very important effects of the curriculum are not being evaluated by achievement tests. Indeed, he observes: "I do not think it is possible yet, and certainly not on the basis of test scores alone, to evaluate whether the Project is successful or not".

Yet another form of evaluation (inconclusive because it was done on a small sample in one school) was carried out by a postgraduate student at UWI, Trinidad. Ishmael[3] showed that more children are continuing science after being exposed to WISCIP, their GCE performance is better and their WISCIP grades correlate well with their GCE grades. These are considerations which are important to teachers in the Caribbean.

The author is currently engaged in a survey in the Caribbean to determine the extent to which the innovations are being used, several years after the official end of the projects. It appears, just before any published materials have become available, that in the small eastern Caribbean countries (where a dissemination base remains for WISC) most schools are using WISC; otherwise the variety of curricula being used in schools is tremendous.

In a report to the Jamaica Curriculum Committee and to CRSP, Reay and Turner[4] analysed the content and treatment of major concepts in MONA and WISCIP/C, this being an intrinsic evaluation of the curricula.

In a familiarization exercise for postgraduate students at UWI Trinidad, students were asked to make a comparison of the units of Trinidad WISCIP and WISCIP/C. It is almost certain that the curricula are uneven from one unit to another, depending on the consultants involved in them, but another determinant of preference appeared to be the degree of authoritarianism of the students (who were all experienced graduate teachers).

A more careful appraisal of the WISCIPs, but not a comparative one, was made by teacher-educators at a workshop in which they developed a check-list for the evaluation of integrated science curricula, and applied it to whichever WISCIP they knew best[5].

[1] Adey, Reay and Turner, The Caribbean ..., op. cit.

[2] Williams, Report on the ..., op. cit. and Report to Institute ..., op. cit.

[3] M. Ishmael, A Summative Evaluation of West Indian Science Curriculum Innovation Project (WISCIP) at Naparima Girls' High School, Trinidad, UWI, 1975 (Unpublished).

[4] J.F. Reay and A.D. Turner, The Teaching of Science in Jamaica, Grades 7-9. Report to Jamaica National Curriculum Science sub-committee and to CEDO/UWI Caribbean Regional Science Project, based on trials of integrated science curricula, Jamaica, UWI, 1973, (mimeo).

[5] J.F. Reay (ed), "Project Appraisal", in Integrated Science and Teacher Education, Trinidad, UWI/Unesco, 1974.

Conclusions

Beeby[1] has said "In my experience, the average teacher has a very great capacity for going on doing the same thing under a different name". The teacher, given a curriculum guide on <u>The Gases of the Atmosphere</u> (O_2, CO_2, N_2, etc.), is very likely to add all he knows about H_2S, SO_2 and perhaps other gases. Evaluation often presumes innocently that the teachers will use the curriculum guides as was intended by the headquarters team. Many of them will do so; but a significant number of these will drop the curriculum after one or two years. The rest will never have taught the curriculum as was intended.

Did the information provided by the evaluators help in the decision-making process? Probably not. Teachers seemed to base their decisions on whether they liked the look of the guides or not. It appeared that the other decision-makers did not find any of the information compelling enough; and in the end, Jamaica's curriculum committee made no choice at all. The most important information sought by the decision-makers appeared to be the kind that would be provided by publishers - would they publish and when would the publications appear? The availability of books for children seemed more interesting than anything else and the Caribbean evaluations apparently provided only academic interest.

Would different strategies have proved any more useful? It is doubtful whether different techniques, or evaluation of different outcomes, would have fared any better. It is possible that a greater involvement of teachers, as associations, would have attracted greater attention to the evaluation results. At that time, the science teachers' associations of the region were relatively weak; now, with more of them, a regional association and with national and regional associations compelling respect, it is possible that an evaluation based on close consultation and assistance from the associations would have an impact. One still feels, however, that the information which teachers will want, will not be of the statistical kind obtained from children but data on how many teachers like it and when the books will be ready. The determined individuality of teachers is something curriculum developers often fail to recognize.

There is no doubt that sound decision-making should be based on valid and reliable evaluation data; perhaps teacher education should pay more attention to convincing teachers of this and should help them evaluate the data they receive.

[1] C.E. Beeby, "Curriculum planning", in Howson, G. (ed), Developing a New Curriculum, (p. 39-52), London, Heinemann, 1970.

ANNOTATED BIBLIOGRAPHY

Beeby, C.E. Curriculum planning. In: Howson, G. (ed), Developing a New Curriculum, (p. 39-52), London, Heinemann, 1970.

A sympathetic treatment of the problems of innovation, especially in developing countries.

Reay, J.F. (ed), Integrated Science and Teacher Education. Trinidad, UWI/Unesco, 1974, p. 237, appendices. Report of Unesco/CEDO workshop in the Caribbean.

Pages 35-43 contain a check-list for project appraisal, developed at the workshop, and report on its application to the WISCIPs. Pages 27-33 summarize a report on the Caribbean summative evaluation, and participants' reaction.

Williams, I.W. WISCIP: A Case Study in Curriculum Innovation, London, Commonwealth Secretariat, 1973, 71 p., appendices.

This report was commissioned by the Commonwealth Secretariat and has subsequently been brought up to date. The Secretariat hopes to publish an abridged version, but may be able to make the full text available to interested individuals. Curriculum developers with small resources will find it most useful.

12 Case Study A: Evaluation of the Scottish Integrated Science Syllabus

A. W. Jeffrey
Her Majesty's Inspectorate, Scotland

SUMMARY

This case study outlines the development and evaluation of a course of study in integrated science for the first two years of secondary education in Scotland. It describes how the principle of "rolling" development is carried out, outlines the various evaluative procedures which were used and gives some of the results of these procedures and the conclusions which the authors of the syllabus drew from them. With little money to spend at the beginning, the group had to use whatever sources of help were available within the system at little or no cost. Some of the procedures used and the validity of some of the conclusions drawn might not meet the stringent appraisal of professional researchers. Yet the group was satisfied that it was getting sufficient evidence to support its belief in the suitability of the course of study which it produced and its contention that pupils of all abilities could learn adequately in following it. The final section indicates the nature of the continuing investigation and some of the more recent developments in consequence.

Science education in Scotland traditionally begins with two years of general science before splitting up into the separate disciplines for presentation in the external examinations. With the success of alternative syllabuses introduced in 1962 in 'O' grade physics and chemistry for academic pupils, it became desirable to produce a generalized approach to science for the early secondary years using similar content, equipment and methods. A working party was set up for this purpose by the Secretary of State for Scotland in December 1964.

Shortly after this, a national decision was taken to restructure secondary education in Scotland along comprehensive lines. In consequence, the working party was asked to consider the nature of science education in the first two years of comprehensive school. It should be stressed that this project was developed mainly by voluntary, unpaid helpers, much of it in their own time. Working within a small budget, it was not possible to undertake many of the standard evaluation procedures normal today. The report containing a new syllabus was published in 1969 as Curriculum Paper 7 - "Science for General Education"[1]. It is the evaluation of this syllabus and its accompanying teaching materials which is the subject of this case study.

Scotland, with a population of about 5 million, has 450 secondary schools. These are the responsibility of local authorities and the curriculum within each school is the responsibility of its headmaster and staff. Nevertheless, curriculum development frequently takes place at a national level and is seen as continuous, rather than as project-based. Programmes are kept under continuous review and course modification is introduced, hopefully, at a rate which teachers are prepared to accept. As a result, evaluation is also continuing sequences of reflective, formative and summative steps.

[1] Scottish Education Department, Curriculum Paper 7: Science for General Education, Edinburgh, HMSO, 1969.

Reflective evaluation

During 1965, general aims were stated and restated. Course content was produced in parallel and judged on its potential to fulfil these general aims. The traditional lower secondary science course had been totally content-based but the working party decided to seek more process-based material which would at the same time deal with concepts relevant to everyday life and be within the pupils' comprehension.

Three major constraints affected decisions. Problems of cost dictated that the content selected should be accomplished with laboratory equipment known to be already in schools. The Examinations Board has to be satisfied that the content used could be examinable at 'O' grade (that is, an external examination administered at 16 or over after four years of secondary education) and was a suitable base from which their examination syllabuses in physics, chemistry and biology could be taught in the succeeding two years. This and the knowledge that not all teachers would be willing to teach integrated science required a fairly even balance of physics, chemistry and biology. During this time, evaluation was carried out by working party members and subject panels of the Examination Board on a subjective basis. The chosen content was finally manipulated into a form suitable for trial.

Using this content, a set of summative objectives, including some affective objectives, was prepared and evaluated and restructured by the working party. It was decided to use the language of Bloom's Taxonomy[1] in stating these. In 1965, this was the first attempt ever made in Scotland to indicate that education in science might be concerned with attitudes, and in hindsight the working party was probably too timid, although it seemed at the time a very courageous step.

Formative evaluation

In writing sections, authors tried the material with their own classes and re-wrote them as a result of the classroom trials. The first full trial began, however, during the 1965-66 session in thirty-six schools, selected for their proximity to one or more members of the working party and thus easy access to advice as well as convenience for evaluation visits.

During this time, feedback was mainly oral obtained at meetings of local groups at two monthly intervals. The points generally discussed included: the reaction of teachers to materials in trial; the reaction of pupils to materials; the problems of presentation and equipment; the success or failure in teaching content and concepts; the time taken to cover work under trial; and the results of internal tests on material taught.

The five HM Inspectors of Schools on the working party were involved in formative evaluation at a more detached level. In their visits to schools, they gained impressions of pupil/teacher, pupil/material and teacher/material interaction. They also tried to evaluate the points listed above to moderate the feedback from teachers. Not all teachers were equally reliable in their evaluation, nor was the information provided by different schools equally well-founded.

As information about the trials of each section was collated, it was rewritten for a second trial. This was repeated twice more and the number of schools increased until 130 were involved in the final trial.

Several important things were discovered. Section 2 of the first draft was Energy and section 3 Electricity - a form of energy. It became clear that this was too soon to tackle electricity and so in the second run it became section 7. It was also discovered that difficulty in constructing circuits was inhibiting the learning in electricity. Circuit boards were introduced to minimize this difficulty. Ohms Law in the original draft was reduced to the more simple $V\uparrow I\uparrow$. A section on space travel was removed because the conceptual difficulties were too great for quite able pupils. In the original, biology was to be taught in the final two of a six-term course. Results were poor and restructuring spread biology more evenly throughout.

[1] Bloom et al, Taxonomy of Educational Objectives, vol. 1 and 2, Longmans, 1956.

It was then also that worksheets were suggested and first drafts produced. Initially they were seen as a guide to the non-specialist teacher indicating sequence, depth of cover and processes to be learned as much as a guide to the pupil. Feedback showed that our estimate of the child's vocabulary was quite wrong and subsequent rewrites tried to reduce it to a more suitable level. It was thought by the time of general publication in 1969 that they were intelligible to all who could read but this has proved to be still too optimistic and they are again being revised.

Another thing which emerged was the lack of skill among teachers in constructing suitable test materials. Because of our use of Bloom's Taxonomy, teachers were familiar with the increasing complexity involved in recall, comprehension, application and so on, but many were unable to write test items to discover whether pupils were learning at these levels. In consequence a national training programme on objective testing was launched among science teachers during the subsequent four years and it was decided to produce and administer centrally-prepared tests. Only a small bank of items could be produced because of cost. Again, with hindsight, this was not good policy but the working party had a very small budget and little could be done at the time to augment it.

One test was produced by Chaplin[1] on the first four sections of the syllabus, with equal numbers of recall, comprehension, application and highest ability items (this was the modified Bloom Taxonomy used by the working party). It was administered to a sample of 1,100 pupils in the trial schools. The verbal reasoning quotient (VRQ) of each pupil in the sample was also obtained from the school, this being the only measure of general ability readily available. The results of the four section test are shown in Table 23.

Table 23. Student results on first four sections of Scottish Syllabus.

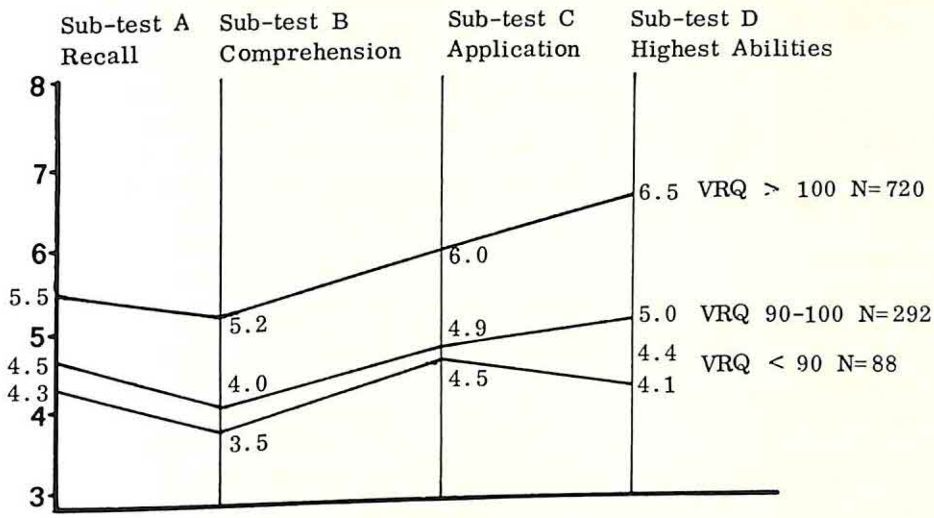
Section tested	Results (expressed as a percentage)
Sub-test A (11 items at level of recall)	46.8
Sub-test B (11 items at level of comprehension)	43.3
Sub-test C (11 items at level of application)	51.8
Sub-test D (11 items at level of highest abilities)	50.8
Over-all mean	48.06

From this it was concluded that pupils were learning adequately at all the levels of complexity for which items had been prepared.

Since the working party was also concerned about whether pupils of all levels or ability were learning adequately, an alternative breakdown of the test results was investigated. Three groups of pupils were identified according to VRQ, (i) > 100; (ii) 90-100, and (iii) < 90. The performance of each group on each sub-test was then determined. The results are shown in Figure 13.

[1] J. Chaplin, unpublished material held by the Scottish Education Department.

Figure 13. Performance of subgroups of pupils with varying verbal reasoning quotients.



It was concluded from this that all levels of ability were learning from the work they were doing. This was encouraging for the success of mixed ability teaching. Various investigations have shown that in Scotland, tests applied to a group of over 500 randomly selected pupils produce results very close to those obtained from national groups. These results could therefore be taken as reasonably indicative of national performance.

A second test constructed on section 7 on electricity was administered to 400 pupils, fifty from each of eight schools. Samples were stratified to represent the different organizations of streamed and unstreamed, selective and unselective groupings. The results indicated that streaming had little effect in improving the performance of the more able but had a considerable and lowering effect on the least able.

Summative evaluation

By late 1968, with 130 schools involved in the trials and the curriculum paper in preparation for publication, it was decided to begin some form of summative evaluation. Money was still in short supply. Chaplin produced a 64-item test on the full first-year course. This was administered to a sample similar in size and composition to that used for his test on section 1-4 and similar results were obtained.

The following session (1969-1970), King[1] undertook to assess the cognitive achievement in the second-year course. He prepared and administered tests on each of the sections 9-15. Because of the difficulty some schools had in finishing the syllabus, the tests for the different sections were not always administered to the same number of students.

King concluded that, although there were now clearer differences in performance between more and less able groups, the learning of all levels of pupils was still reasonable and sufficient to justify retaining the syllabus at the standard then in use. It is doubtful if the evidence really supported these conclusions, and later work has shown that they were too optimistic.

During 1970, there was considerable debate about the inclusion of the affective objectives in the general statement of intent. It was also clear from visits paid to schools by HM Inspectorate that many teachers were not attempting to teach towards attitudinal objectives. In consequence, it was fortunate that larger sums of money became available for the first time to mount a properly funded evaluation programme, and Brown undertook the work, as reported in the next chapter.

[1] J. King, Unpublished M.Ed. Thesis, University of Glasgow, 1970.

Curriculum paper 7 was written as a brief non-technical statement of the place of science education in the curriculum and the form which that education should take in the early years of a comprehensive educational system. It was therefore inadequate as a basis for the more detailed research of Brown. In any similar document in the future, it may be that the theoretical arguments for the structure and objectives of the recommended syllabuses should be included. Hamilton[1] sees other reasons for teachers not following the recommendations of Curriculum paper 7. Brown's study shows that children, however taught, are less enamoured of science after two years of secondary education than they were at the beginning. What can be done about this must exercise all of us for some time to come.

Second phase evaluation

In Scotland, each school head and staff are collectively responsible for the curriculum and consequently for the various syllabuses which form that curriculum. Brown's research included, in 1971-72, an investigation on survey of enrolments in integrated science. It was discovered that, by 1972, somewhere between 65 to 70 per cent of our schools were using it at least in the first year. What was also true and less satisfactory was that not all were teaching to all of the objectives. In consequence, during the 1971-72 session, a survey was varried out by HM Inspectorate. Inspectors visited and assessed the science education in a large sample of schools using the syllabus. Elements of their assessment were included in a report on the state of education in the first two years of secondary education in Scotland[2]. Further consequence of this survey was a decision to review the whole programme of work and such emerging problems as the vocabulary of the work sheets, the need for material more carefully structured for uses in mixed ability classes and the difficulty of establishing certain concepts. An investigation by staff in Dundee College of Education into the comprehension of the structure and behaviour of matter was also initiated. This is a key concept in the syllabus and was found to be causing difficulty with pupils. Early results show that even the most able at age 12-13 do not comprehend the model but further work is in progress in an effort to check this and to discover where the difficulties lie.

As a result of this research and further reflective and formative evaluation, the contents, sequence and scope of the work are being completely restructured, the worksheets produced in a more pictorial form and the whole re-tested in a selection of schools. This time, many more teachers are being involved at the creative stage - some sixty teachers in ten working groups spread across the country - seeking wider diffusion of the objectives of the course. To co-ordinate the work of this group, money has been made available for the appointment of a full-time co-ordinator and for full-scale evaluation. A more detailed teacher's guide is also being produced for each section to establish more specific outcomes for the work and to give teachers more detailed guidelines towards achieving these outcomes.

During 1974-75, trials of this rewrite took place in a small number of schools concentrating on the achievements of the least able. It is believed that if the level of achievement of this group can be established, the extension work for the other groups of greater ability can be more easily determined. After rewriting, as a result of the feedback, each section will be tried in fifth schools during 1975-76 session.

Tests have been produced for each section and the achievement of the various objectives can be monitored for different schools and for different ability levels. This work is being carried out under the direction of Kellington[3].

[1] Hamilton Reid and Walker (ed), Case Studies in Curriculum Change: Great Britain and USA, Routledge, Kegan and Paul, 1975.

[2] HMSO, The First Two Years of Secondary Education, The report of a survey by HM Inspectors of schools in Scotland, Edinburgh, HMSO, 1972.

[3] S. Kellington, work in progress at Notre Dame College of Education, Glasgow (1974 on).

It has frequently been said in Scotland that the teaching of an integrated science programme in the early years inhibits the proper growth of learning in chemistry or physics or biology afterwards. In order to investigate this claim, the performance of candidates at the 'O' grade (fourth year) examinations is monitored from time to time. In 1970 about 20 per cent of pupils could be expected to have experience of the integrated science course; in 1974 it was nearer 70 per cent.

National examinations

A sample of such monitoring is shown in Table 24. The objective tests used at national level have so far been kept confidential and items have been used more than once. A new FV has been calculated for each re-use of each item.

Table 24. National examination results in Scottish integrated science.

	Pre-test			First use			Second use		
Item	Year	FV	r_b	Year	FV	r_b	Year	FV	r_b
1	1972	.75	.56	1975	0.80	0.54			
2	1970	0.68	0.44	1972	0.57	0.58	1975	0.62	0.62
3	1970	0.40	0.43	1971	0.30	0.32	1975	0.34	0.16
4	1970	0.57	0.39	1972	0.61	0.55	1975	0.65	0.61
5	1970	0.60	0.47	1973	0.66	0.54	1975	0.61	0.56
6	1970	0.43	0.62	1971	0.56	0.59	1975	0.52	0.59
7	1970	0.62	0.54	1971	0.66	0.45	1975	0.66	0.52
8	1970	0.53	0.48	1972	0.66	0.62	1975	0.62	0.65
9	1970	0.70	0.57	1972	0.76	0.66	1975	0.70	0.65
10	1970	0.77	0.29	1972	0.76	0.36	1975	0.74	0.42

FV - Facility value
r_b - Point biserial correlation

The facility values show that there has been no change in the general standard of performance in the examination. The significance of this evidence lies in the fact that during the period covered by the investigation, the numbers examined in physics and chemistry have increased by about 50 per cent and this increase contains a large number of candidates previously considered unable to cope with the difficulty level of the syllabuses. It should be noted that during this time the total number of pupils in an age group varied by less than 10 per cent.

During the life of the integrated syllabus other facets have been investigated unofficially by individuals or groups. Hamilton, investigating school-based responses to innovation, used the science departments of two schools trying to introduce the integrated science syllabus. Much useful information has been obtained about the quality of our original communications and about the attenuation of objectives between intention and practice. Roebuck[1], investigating the potential of self-instructional materials in remote schools (PHI) used sections of the Integrated

[1] Roebuck et al, Project PHI Programmed Materials in Remote Secondary Schools in Highlands and Islands of Scotland, University of Glasgow, 1974.

Science syllabus because of the related specific objectives. From PHI a great deal was learned about alternative modes of presentation, including a game which improved comprehension of electrical circuitry.

In this necessarily brief report of a decade of activity, it has only been possible to outline the efforts of numerous dedicated people. Probably some of the things done in the early days would now have been done differently. Others who may try to follow can learn from our mistakes as, certainly, we have done. Most of what we have done appeared reasonable in the circumstances.

Footnote: I would like to acknowledge publicly all those who have given so willingly of their time and their expertise for these activities.

13 Case Study B: Evaluation of attitude objectives

Sally Brown
University of Stirling, Scotland

SUMMARY

This case study exemplifies evaluation of the attitudinal aspect of integrated science education. The approach is from two perspectives: firstly, the purposes, meanings and feasibility of a group of attitude objectives are examined, and secondly, the achievement of these objectives by pupils and the factors influencing that achievement are investigated.

INTRODUCTION

Although there have been increasing emphases on attitude objectives in integrated science courses, examples of strategies appropriate for evaluation of this aspect of the curriculum have not received much attention in the literature. This case study attempts to provide such exemplification.

The study, financed by the Scottish Education Department, centred on five attitude objectives proposed for pupils in the first two years of secondary school in Scotland. The intention was that these pupils should acquire:

A1. awareness of the inter-relationship of the different disciplines of science,

A2. awareness of the relationship of science to other aspects of the (school) curriculum,

A3. awareness of the contribution of science to the social and economic life of the community,

A4. interest and enjoyment in science,

A5. an objectivity in observation and in assessing observations.[1]

The evaluation was conceptualized as two broad questions, the first largely non-empirical, the second empirical:

1. Are the arguments presented for inclusion of the attitude objectives educationally valid, are the meanings and purposes of the objectives clear? To what extent do they correspond to attitudes that children exhibit?

2. To what extent are these objectives achieved by the pupils? What factors influence that achievement?

The non-empirical study

The evaluator's task here was seen as one of 'curriculum criticism' (analagous, perhaps, to literary criticism) whereby the critic attempts to reveal to the audience (e.g. science teachers) the meanings of the various ideas inherent in the curriculum.

[1] Scottish Education Department, op. cit., p. 16.

The arguments for including attitude objectives in the course

The course document CP7 (Curriculum Paper 7, Scottish Education Department, 1969) was examined for justifications for inclusion of these objectives. This was done from three perspectives.

Firstly, a search was made for direct arguments for emphasis on affective goals. In general, such arguments as they occur in the literature can be grouped into seven categories. First, the concept of the ideal society and good citizens: assertions are made about the characteristics of an ideal society, the 'good' citizen's role and appropriate attitude objectives for the production of such individuals. The next category involves desirable social values and behaviours: attitude objectives are selected to produce 'good-citizen' qualities for society as it is, rather than for the ideal. Pupils' social adjustment regarded objectives as a means towards fulfilment of pupils' needs for social adjustment. Another category was rational, intelligent, valuing behaviour: attitude objectives are instrumental to the achievement of such behaviour. Attitudes conducive to cognitive functioning were yet another affective goal; attitudes (e.g. objectivity) are valuable in cognitive tasks of scientific enquiry. Then there were affective routes to cognitive goals: for example, interest and satisfaction are means to improved cognitive performance. Finally, meeting pupil needs (other than social adjustment): objectives relate to pupil needs for satisfaction, enjoyment, etc.

CP7 provides no such arguments for inclusion of its general attitude objectives. However, the nature of the specific objectives for each section of work perhaps implies arguments of types involving cognitive functioning and cognitive goals.

The second approach looked at the particular meanings ascribed to 'integration' by this curriculum, and those attitudes logically related to such meanings. The meanings of 'integration' may be classified into three groups. Firstly, unity of all knowledge: knowledge is seen as one conceptual structure and/or enquiry process. Secondly, unity of the disciplines of science: science is seen as one conceptual structure and/or enquiry process distinct from others and without component disciplines. Thirdly, interdisciplinary study: science is seen as the study of topics from viewpoints of various disciplines.

Integration in the first sense would lead us to expect any development of scientific knowledge to affect all areas of our experience. This would imply attitude objectives which reflect knowledge of and feeling towards the pervading influence of science on our culture, ourselves and our environment and toward science as part of the unified structure of knowledge. The logical attitudinal implications of the second meaning are limited to the inter-relationships among the disciplines of science. 'Interdisciplinary study' implies no particular attitudes towards science other than as a collection of established disciplines.

Each of these meanings of integration have been used in CP7, though there is some lack of correspondence between the rhetoric of the document and the curriculum materials in that the organization of content in the latter reflects only 'interdisciplinary study'. In so far as the first sense is accepted, there is justification for inclusion of the first three attitude objectives (A1, A2, A3.)

The third possible source of justifications was the rationale for the integrated course itself. Arguments for presenting science in an integrated form are generally in terms of anticipated outcomes from such a course or the constraints and demands under which the course must be implemented. The extent to which such outcomes might be expected to emphasize attitudes and such constraints and demands to constrain or demand achievement of attitude objectives will have implications for inclusion of such objectives.

The types of arguments found in the literature for the value of an integrated approach may be grouped in the following six categories:

1. Outcomes demanded by society, e.g. provision of scientists, informed lay population, informed political leadership.

2. Resource constraints, e.g. accommodation, equipment, time, teachers.

3. Political demands, e.g. common-core course for all pupils, national assessment system.

4. Conditions for effective learning, e.g. pupil security, motivation, interest.

5. Conditions for effective teaching, e.g. teachers' interests, competence.

6. Demands imposed by the subject, e.g. unified nature of scientific enquiry.

Types 2 and 5 (concerned with resource constraints and advantages for teachers) do not imply attitude objectives for pupils. Type 1 suggests attitude objectives concerned with the broad influences of science on society; type 3 might (if the 'demand' were 'mixed ability groups following a common course') dictate the development of attitudes that imply science as involving the whole community, type 4 may imply particular attitudes (e.g. interest) seen as related to pupil learning; type 6 will imply attitudes related to the particular meaning of 'integration' used, and those attitudes (such as objectivity and scepticism) assumed to be characteristic of scientists.

CP7 provides arguments for 'integration' of type 1 implying objective A3 (social and economic implications), type 3 implying objective A3, type 4 implying objective A4 (interest and enjoyment) and type 6 implying objective A5 (objectivity) and, in so far as 'integration' means the 'unity of all knowledge', A1 (inter-relationships among sciences), A2 (relationship of science to other subjects) and A3.

Clarity of the objectives

CP7 provides no theoretical framework to clarify the purposes and meanings of its objectives.
The objectives were therefore compared with a framework mapped out by the subject-independent
categories of the Affective Taxonomy[1] together with Klopfer's[2] categories of affective behaviours
in science. Figure 14 outlines the broad correspondence between these two groups of categories.
Comparison with this framework identified those objectives for which further specification or
clarification seemed necessary.

The first three objectives were readily related to Krathwohl's lowest level 'Awareness'.
However, lack of description of the cognitive behaviours that the writers had in mind for these
objectives prevented unambiguous location on Klopfer's dimension. For example, there was no
indication within the course materials about what sorts of relationships between the sciences the
pupils were to become aware of.

A4 ('interest and enjoyment') covers three Klopfer (H4, H5 and H6) and three Krathwohl
categories. This suggests that closer specification of the level of affective behaviour is required.

A5 ('objectivity') is identified with Klopfer's H3 category and so covers three Krathwohl levels.
It is not clear what level of commitment to 'scientific attitudes' is expected from this age-group.

It is surprising that CP7 has no referents for Klopfer's first three categories. Nevertheless,
if that gap were filled, necessary cognitive and affective behaviours specified and a level of
commitment identified, these affective objectives could form a comprehensive list which might be
related more readily to classroom procedures.

Relationships of attitude objectives to pupils' attitudes

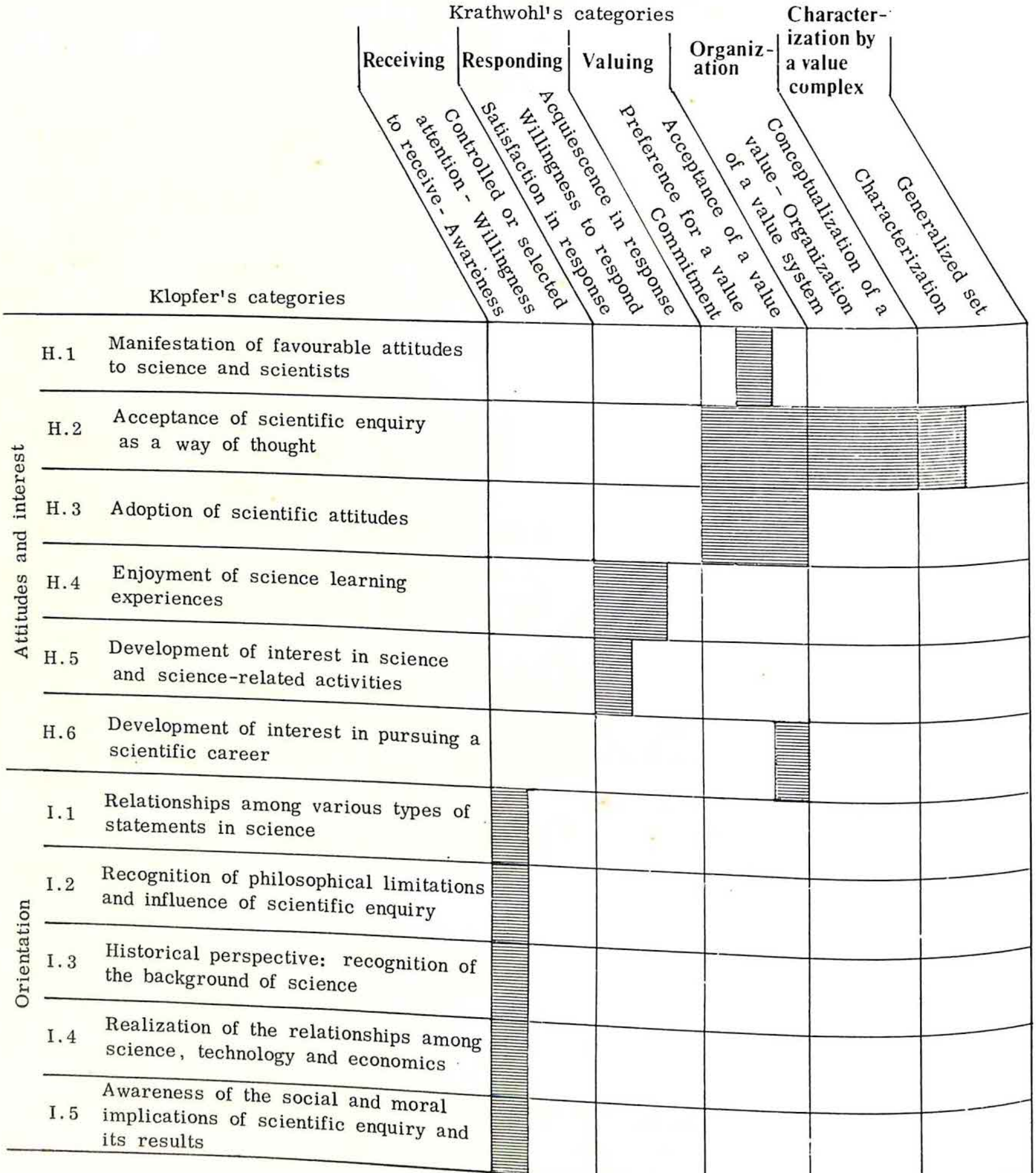
To what extent do these objectives correspond to independent attitude dimensions exhibited by pupils?
Five attitude scales corresponding to the objectives were developed[3]. Items for these scales were
collected from conversations with pupils and were submitted to eight 'experts' who judged whether
or not each item related (positively or negatively) to one of the five objectives. 'Position

[1] D.R. Krathwohl, B.S. Bloom and B.B. Masia, Taxonomy of Educational Objectives Handbook II: Affective Domain, London, Longmans, 1964.

[2] Klopfer, "Evaluation of Learning ...", op. cit.

[3] S.A. Brown and T.N. Davis "The Development of an Attitude to Science Scale for 12 to 14 year olds", Scottish Educational Studies, vol. 5, no. 2, 1973, p. 85-94.

Figure 14. Relationship between Klopfer's categories of affective behaviours in science education and Krathwohl's taxonomy of affective objectives.



(Shaded areas indicate the level of behaviour in Krathwohl's hierarchical system that corresponds to each of Klopfer's science categories.)

statements' made by the authors of CP7 and providing expansion and clarification of the objectives were used by the 'experts' in their judgements. The items were then piloted and relationships between pupils' scale scores and other variables (e.g. age and sex) were hypothesized and tested, resulting in the establishment of some degree of construct validity alongside content validity for the scales.

The five scales were then administered to 2,815 thirteen-year-old pupils. The responses were factor analysed which, speaking approximately, clusters together those items that pupils tend to answer in the same way, i.e. by 'agreeing' or 'disagreeing' with them. If the attitudes displayed by pupils' responses correspond exactly to the objectives, then we would expect each 'cluster' of items to correspond exactly to each 'scale'.

The results suggested that all the objectives except A3 (social and economic implications of science) correspond closely to distinct pupil attitudes. A3 had no clear referent, but the negative items for this scale formed an 'Alienation from Science' factor. However, only 36 per cent of the variation in responses was accounted for by these factors, suggesting that either the pupils' responses were predominantly idiosyncratic, or there are further attitude dimensions that are untapped by substantial numbers of items from these scales.

The empirical study

In the absence of a theoretical framework, the research questions, hypotheses and design were related to questions and hypotheses identified by teachers and inspectors, empirical evidence from other research and preliminary pilot work.

This was unsatisfactory for several reasons. The selection of 'important' issues and 'information' supplied by teachers and inspectors appeared idiosyncratic and often conflicting. Small studies of attitudes to science have been numerous but largely concerned with 'interest' and unsuccessful in establishing relationships between attitudes and other social and personal variables. Relationships appearing in pilot studies are notorious for disappearing in larger-scale work[1]. Hypotheses formulated on such bases are tentative and very different in nature from scientific hypotheses based on theory.

Three categories of research questions were formulated. What are the differences in attitude achievement (i) among schools (a) all following integrated science and (b) following various courses, (ii) among pupils within schools and (iii) among pupils within teaching groups?

A sample of 2,815 pupils from forty schools stratified by size, location, denomination and type of science course was taken for the 'between-schools' study. Twelve of these schools formed the 'within-schools' sample and thirty-eight teaching groups from these twelve schools formed the 'within-teaching-group' sample.

Attitude scales relating to the five objectives were administered on entry to secondary school (twelve-year-olds) and at the end of second year.

Between-schools study

A regression analysis on the five mean attitude scores for boys and girls separately from each school was used to determine the variation in attitudes that could be accounted for by differences in characteristics of the groups of pupils (sex, social class, IQ, divergency, attitudes on entry to school), the schools (size, location, denomination), the science course (integrated or separate sciences), and the science class (size, ability and sex grouping, time allocation).

About half of the variation (44 to 62 per cent) in scores on each scale at the end of second year was accounted for by characteristics of pupils assessed on entry to school. School characteristics showed no significant effect on pupils' attitudes except A2 (16 per cent additional variance accounted for, small schools outside the cities being more aware of the relationship of

[1] J. Raven, J. Ritchie and D. Baxter, "Factor Analysis and Cluster Analysis: Their Value and Stability in Social Survey Research", Economic and Social Review, vol. 2, no. 3, 1971, p. 367-92.

science to other subjects). The type of science course followed accounted for significant additional variance (11 per cent) in A1 scores only, and pupils following separate sciences appeared more aware of the inter-relationships among the sciences. This may well be a direct reflection of the failure of the integrated course to make clear the nature of these inter-relationships. Characteristics of the science classes influenced A5 scores only (6 per cent, mixed sex, but not mixed ability, groups with a larger number of periods of science displaying higher 'objectivity').

All the scales except A4 ('interest and enjoyment') showed significant increases in scores over the two-year period for all groups. These correspond to objectives with substantial cognitive components. A4 which is essentially affective showed a decrease for the whole population.

Within-schools study

The relationships between individual pupil's attitude scores and the other pupil measures were explored using regression analysis and patterns of relationships common to schools following the integrated course were searched for.

No such patterns were found for this or any other group and the results of all twelve schools were combined into one regression. The patterns of variance accounted for by pupil characteristics were similar to those of the 'between-schools' but less substantial (16 to 29 per cent, cf. 44 to 62 per cent).

Within-teaching-groups study

Three relationships considered were those between a pupil's attitude score at the end of the second year and school science achievement, teacher's rating of 'academic ability', and teacher's rating of 'interest in science'.

Median correlations between attitudes and achievement were only moderate (0.32 to 0.36). The relationships between attitude scores and teachers' 'academic ability' ratings were similar, but there was some indication that the pupils' attitudes had shifted into line with the teachers' expectations of them during the two years. (Other research[1] has suggested that pupils' performances on cognitive tasks are influenced by their teachers' expectations of them and that the behaviour is in accordance with the expectation.)

Teachers' ratings of pupils' 'interest' were even less strongly related to pupils' attitudes ($r = 0.29$ to 0.39), but correlated 0.82 with pupils' achievement scores. This suggests that the teachers based their assessment of pupils' interest largely on academic performance. This apparent failure to recognize behaviours reflecting levels of interest suggests that the teachers may not be consciously teaching towards interest goals.

Concluding remarks

With hindsight, selection of variables on 'intuition' that 'schools that are large, in inner-city areas, giving few periods to science, following (or not following) integrated courses, result in unfavourable attitudes to science', was, in this context, an unwise procedure. These factors per se exert little influence on attitudes in contrast to factors outside the secondary education system's control, i.e. characteristics that groups of pupils have on entry to school (though the effect of the latter is much reduced if individual's scores are considered).

If we seek educational variables that are useful predictors of the still unaccounted for variation in attitudes, it appears we must abandon the demographic and organizational type of variable, develop a theoretical framework for the study of attitude achievement (using available psychological information), and, for fruitful variables, look to patterns of teacher-pupil communication, teachers' expectations of pupils, topics covered in the course and teachers' classroom tactics and strategies.

[1] W.V. Beez, "Influence of Biased Psychological Reports on Teacher Behaviour and Pupil Performance", in A. Morrison and D. McIntyre, Social Psychology of Teaching, Harmondsworth, Middlesex, Penguin Books, 1968.

ANNOTATED BIBLIOGRAPHY

Brown, S.A., Affective Objectives in an Integrated Curriculum, Ph.D. thesis, University of Stirling, 1975.

This provides a full account of the evaluation including a detailed examination of the information available from social psychology about the ways in which attitude goals may be achieved.

Klopfer, L.E., "Evaluation of Learning in Science", in Bloom, B.S., Hastings, J.T. and Madaus, G.F., Handbook on Formative and Summative Evaluation of Student Learning, p. 559-641, London, McGraw Hill, 1971.

A classification scheme for affective behaviours exhibited by science pupils. Emphasizes attitudes that relate to science as a system of enquiry and carefully avoids ambiguity by clear delineation of behaviours.

Krathwohl, D.R., Bloom, B.S. and Masia, B.B., Taxonomy of Educational Objectives Handbook II: Affective Domain, London, Longmans, 1964.

A heirarchical classification of affective objectives that has greatly increased precision in this area and improved communication among curriculum developers and teachers.

Oppenheim, A.N., Questionnaire Design and Attitude Measurement, London, Heinemann, 1966.

A comprehensive review of, and introduction to, attitude measurement.

Willis, G.H., "Curriculum Criticism and Literary Criticism", Journal of Curriculum Studies, vol. 7, no. 1, 1975, p. 3-17.

Curriculum criticism is seen as an aesthetic aspect of curriculum enquiry that is concerned with the question 'How well does this curriculum represent those ideas with which it is concerned?'.

14 Evaluation of the integrated science project in Brazil

Myriam Krasilchik
Sao Paulo, Brazil

SUMMARY

This paper describes the designing of the evaluation component of the Integrated Science Project developed in Brazil; it discusses decisions, points and problems faced by a project director when planning and conducting an evaluation study. Expected and unexpected results are outlined as well as possible alternatives to solve problems in similar situations. Evaluation of this project is not yet completed and thus this report is preliminary rather than final in nature.

INTRODUCTION

A brief outline of the present situation of integrated science in the Brazilian school system is a necessary pre-requisite to an analysis of the evaluation of integrated science teaching in Brazil.

The Brazilian school system is composed of a compulsory Fundamental school lasting eight years followed by a High school lasting three years which aims at giving pupils occupational training as well as preparing them for university.

Traditionally, science is taught at Fundamental school as a single subject and at High school as three separate subjects: physics, chemistry and biology. Though general science is part of the school curriculum, this is not integrated science as defined today.

Each Brazilian state has an educational system and structure of its own and prepares its curriculum according to the general recommendation from the Federal Council of Education. Thus the situation varies slightly from state to state.

Usually the scientific content is taught in scattered topics which do not present the unity of science and scientific procedures. Clear examples of this can be found in the publication "Curricula of Basic School Teaching" [1] (a descriptive analysis of the federal proposals for curricula prepared by the Departamento de Ensino Fundamental from the Ministry of Education and Culture). In its analysis of the science curricula, it is possible to see from the curriculum proposals of the various educational systems of the Brazilian states that the most frequent form of content integration combines the so-called physical and biological sciences, dividing them into topics such as living things, matter, energy, earth and the universe, emphasizing aspects of physics, chemistry or biology. The concern for the quantity of information to be transmitted also leads the organizers of the curriculum proposals to a gap between the intended objectives, the recommended methodology and the topics selected for the school programmes.

[1] MEC/DEF - Curriculo de Ensino de 1º grau - Analise descritiva das propostas curriculares das Unidades Federadas - MG, 1973, (an official publication obtainable from the Ministry of Education) - Ministerio da Educacao e Cultura, Esplanada dos Ministerios, Brasilia (DF) Brasil.

Grobman's[1] concept of a curriculum project as involving "group effort to produce some new kind of approach using experimental tryouts of preliminary materials and feedback from such tryouts to be used for the improvement of the curriculum, prior to its release for general distribution", is not common in Brazil.

Innovation by the science teaching movement followed the Sputnik era. Since 1954 a group of Brazilian educators and scientists[2] have worked constantly to improve science teaching. They began with the preparation of original teaching material and later included adaptations of many curricula developed during the sixties in the United States and the United Kingdom. In 1964, the "Initiation to Science" project prepared by the Brazilian Education Science and Culture Institute (IBECC) had several characteristics of integrated science. However, the work was the object of neither formal evaluation nor of regular revisions.

In 1965, the Ministry of Education assisted the improvement of science teaching by establishing Science Teacher Training Centres in the six biggest Brazilian cities. Several different groups in different regions were formed and they worked on the preparation of teaching materials and the development of teacher-training programmes.

Concurrently, a new educational research field evolved on the international scene, namely, curriculum preparation and evaluation, with the consequent accumulation of instruments, experience and data about the process of curriculum development. However, because of a shortage of financial and, even more so, human resources, the Brazilian groups did not utilize the knowledge and methodology developed. A new effort to improve curriculum development and evaluation came from the federal government's organization of the "Science Teaching Project" which sponsored the preparation of a series of projects in different institutions (e.g. science teaching centres, universities and research institutes).

In the present phase of Brazilian education, one may identify several dilemmas and gaps; a nationalist movement searching for its own cultural expression and a marked exterior cultural and technological influence; the urgent need for change and an instability of values; an elitist education and large groups of the population which are economic, social and educational outcasts. Such factors obviously have a major influence on the work of integrated science curriculum research which aims at answering the demand of the Brazilian student population while simultaneously presenting genuinely national and universal characteristics.

The need for curriculum evaluation was a natural consequence for the groups involved in curriculum development, particularly those of integrated science.

However, several obstacles were encountered. The measurement of student achievement is the most common concept of evaluation among administrators. Therefore, it was necessary to demonstrate that the financial investment in evaluation was warranted in that quality control analysis would lead to a greater development of education's potential. It was also necessary to justify the idea that the money spent in the evaluation of instruction materials would lead to greater efficiency in their use, so that teaching and learning could really attain the proposed aims. The efforts of specialists, the broad bibliography on the subject and the need for evaluation because of the failure of many projects are slowly changing the situation. In integrated science, evaluation costs amounted to some 30 per cent of development costs.

Obtaining financial resources did not solve all the problems. The shortage of personnel makes educational research in Brazil still incoherent. The examination of educational research in Brazil[3] from 1968 to 1973 shows that only 7.21 per cent of the work refers to curriculum research. In the work carried out in 1973, the most frequently used technique (48.8 per cent)

[1] H. Grobman, Developmental Curriculum Projects, Peacock, 1970, p. 4.

[2] I. Raw, An effort to improve Science Education in Brazil, (mimeographed), 1970.

[3] INEP - Cadastro de Pesquisas Educacionais no Brasil, 1968/1973, MEC, 1975, (an official publication obtainable from the Ministry of Education) Ministerio da Educacao e Cultura, Esplanada dos Ministerios, Brasilia (DF) Brasil.

was the survey, followed by the interview. The most common treatment is percentage distribution analysis. Factor analysis, the use of scales and multivariate analysis are rare. In the field of curriculum and teaching programme research, hand calculators are more frequently used than computers. The Offices of Education seldom worry about teaching evaluation due to the number of problems, mainly political, which they must solve.

An attempt was made to develop curriculum evaluation by recruiting social scientists and pedagogues to start the work. Foreign consultants came to Brazil to help the initial teams. Institutions such as The Ford Foundation, the Organization of American States (OAS) and Unesco helped Brazilian institutions to systematize and rationalize the field of curriculum projects production and evaluation.

Still, there are other practical problems: access to computer centres is difficult, there are problems in instrument application and the communication system is inefficient in some parts of the country.

Research groups still work very much in isolation. Meetings of educational researchers to discuss the results of their work are rare, as is the publication of specialized journals and periodicals which would allow for an exchange of ideas and a critical analysis of work, thereby improving quality.

National technology is poor and several methods and techniques cannot be used because of the difficulty in obtaining necessary equipment, (i.e. video-tape classroom observation, which could be useful to many projects).

Against the above background, this paper will describe problems faced by a project director organizing an evaluation project within the limits of local conditions.

Evaluation of the integrated science project

Some high schools offer only one science course, with three classes a week. Students normally have a course on one scientific subject (physics, chemistry or biology) depending on the teacher available. These courses emphasize content, are bookish and do not deal with technological or social aspects of science. The schools are poor and laboratories and practical work scarce.

Considering such a situation the Sao Paulo Teachers Training Centre (CECISP) proposed to develop an integrated science project for students in the first year of high school who were not aiming at scientific careers. The project had three major goals: to present basic scientific knowledge, to develop student problem-solving abilities and to help students understand the social implications of science and the complexity of the man/environment relation.

The difficulty of the enterprise was recognized from the beginning. Teachers had not only to use an unconventional text and a new methodology suitable to the project but also consider the social implications of the development of science. This resulted in resistance to the project on the part of teachers who were accustomed to more conventional courses. In addition, the production team for the "developmental project" was composed of people with differing backgrounds, i.e. biologists, physicists, chemists, geographers and anthropologists, producing group interaction that was extremely stimulating, although not always easy. The organization of a curriculum planning team presented some difficulties in terms of creating an harmonious, democratic climate combined with precise allocation of tasks. The inclusion of evaluators in the team was determined from the beginning of the work, but administrative constraints made it impossible to incorporate them regularly.

Development phases of the curriculum project

There were four phases in setting up the project: pre-planning, i.e. preparing a proposal for financing (1972); project planning, to specify goals and aims and to select the contents for development by the project (1973); preparing material and field testing of some chapters and activities and organizing the evaluation study (1974); and experimental application and implementation, involving teacher-training, developing the evaluation study and beginning revision (1975).

Budget distribution by the beginning of the school year had a pronounced influence on the organization of the phases.

The schedule for producing the experimental edition of the integrated science materials assumed that the project would end in November 1974. The school year begins in March in most Brazilian states so evaluation suffered as a result of the need to prepare the experimental edition of the book and the evaluation project at the same time.

In preparing an evaluation project, it appears important to allow a gap of at least six months between the end of the developmental phase and the beginning of experimental application.

Preparation of an evaluation project

The first stage of evaluation involves negotiations and constant communication to acquaint the institution that prepares the material, the financing agencies and the evaluating agencies. From the institutional point of view, the achievement of mutual confidence and harmony at this stage is essential in order to assure a good development of the project.

The project director faces a series of decisions: What are the main questions that should be asked? What is the best "design" for the study to help in the revision of the project? What is the adequate composition of the sample? How should the evaluation team be organized? What are the competencies needed? What should be the relationship between the evaluation group and the development team? How should the study be conducted, and what are the kind of "unexpected facts" that should be "expected"?

If we consider that the integrated science project has a national scope, evaluation should include several regions of the country, a difficult problem given the size of Brazil. Taking into account all the differences found among places, it would be difficult and very expensive to try to select a representative sample. This situation was solved by using the Science Teaching Centres located in six states; they were in charge of the supervision of the regional evaluation. According to the financing agency, evaluation in six states would have been too expensive. It was therefore decided to carry on the study in the Sao Paulo, Recife and Belo Horizonte centres only.

Another decision was to analyse the effects of both the materials and the teacher-training programme. As the Brazilian educational system lacks financial resources, the training required to guarantee efficiency of the project would have been expensive and thus undesirable. Therefore, one of the working hypotheses was that the project could be well-managed with good results even by teachers who had not attended specific training courses.

There was also the question of whether or not to include a group which would not use the materials but which would take the same tests as the experimental groups. Since it would not affect either expenses or the workload by much, a group of this nature was included in the project design.

The main questions to be answered by the evaluation were:

1. Are there detectable differences between pupils who had trained teachers and those who had non-trained teachers?
2. Does the project cause a change in the attitudes of students and teachers towards science?
3. What is the relation between the methodology used and the results obtained?
4. What content does the pupil learn? What are the results in the different levels of the cognitive domain?
5. What should be changed in the project?

Six variables were considered in collecting the data: age and sex of pupils; time of classes, day or evening; grade, final year of Fundamental school or first year of High school; state, place of use; teachers' previous training; and use or not of the project.

Instruments

A test based on the objectives proposed for six chapters of the project. The questions were related to the objectives in the cognitive domain as classified by Bloom[1]. Thus, the test used was composed of 155 multiple choice items with four alternatives, prepared by the authors of the project. The complete test was applied at the beginning of the study and the questions for each chapter reapplied again at the end of respective chapters during the school year.

Attitude scale aimed at measuring the influence of the project on pupils' and teachers' attitudes towards science and implications for social problems.

Questionnaire to check pupil interest and difficulties - consisting of a series of questions with open answers.

Reports prepared by the teachers showing the difficulties they found and suggestions for changes.

Organization

In January 1975, a summer course was organized for sixty teachers from the state of Sao Paulo and ten teachers from other states, all chosen and sent by the Science Teaching Centres. These teachers, who would become regional supervisors, participated in the training course and later stayed for a week at the project head office talking over problems referring to curriculum application and the organization of the evaluation with the team. During this period the evaluation team instructed them concerning the application of the instruments, the organization of evaluation groups, the teaching of training courses for teachers for the experimental team and communication between the evaluation sub-groups and the head office.

Problems of recruiting teachers to use the project coupled with the differences in the beginning of the school year in each state have imposed variations in teacher-training. In some Centres, the course was taught at fixed intervals, in others, it was taught after classes had begun. Though this difference was neither foreseen nor desirable, the variation in training and its possible effects will be taken in consideration in the final evaluation analysis.

The participants in the Sao Paulo training programme were attracted by the opportunity to learn new techniques and by the prestige and status attached to the project.

Controversy also arose about the most appropriate stage at which to use the project materials (the end of Fundamental school or the beginning of High school). Thus, the decision was made to include some pre-high school classes. The final design is shown in Table 25. The study started with 926 pupils and twenty-six teachers.

Table 25. Integrated science project design.

Instrument	Application	Sample
Test	before and after	all pupils
Attitude scale	before and after	pupils and teachers
Interest questionnaire	after	pupils who used the project
Report	after	teachers who used the project

[1] Bloom et al, Taxonomy of ..., op. cit.

Recruitment of teachers

Despite the initial enthusiasm, it was very difficult to find teachers willing to use the project. From the start, there was noticeable resistance. As the curriculum does not aim at preparing the students for university entrance examinations - very important in Brazilian educational life and a strong influence in the high school curricula - teachers and students were opposed to the project. Some private schools, for instance, refused to co-operate in the evaluation for they depend on the influence of the students and cannot risk giving a course that does not meet university entry requirements.

At one evaluation centre, the local supervisors found it necessary to use the force of the state official system to impose the project materials. This led to some disagreeable situations and unwillingness on the part of the teachers.

We believe that the participation in a project must be voluntary, not compulsory. Although this seems obvious, it may be a useful reminder for those who are going to start a project in centralized government schemes.

Another course of resistance derives from teachers who feel insecure teaching a new project. In the specific case of integrated science, combining as it does contents, emphases and methodologies from many fields, such a factor is very important. In these cases, support groups formed by the evaluation supervisors are useful to help the teachers overcome problems.

In the school seen as an institution, there were several different problems among the teachers
of scientific subjects. The introduction of integrated science created competition and worries about
job possibilities. On the other hand, pedagogues and educators who face science teachers involved
in a project of educational research may see it as an invasion of their areas and start to defend
their territories. In some cases, this can be advantageous, resulting in interaction between
specialists and non-specialists.

The present legislation does not refer to integrated science, another obstacle to the adoption
of the project. Many teachers and administrators prefer to remain with the traditional school
subjects rather than enjoy the freedom which the law allows them.

Although the integrated science project demands only a minimum of very inexpensive equipment,
the difference between the budget and the access to the laboratory material in different regions may
also represent source of opposition resistance points.

The leadership of the supervisors, their available time and their prestige in the community,
are also factors which influence adoption. Whenever possible, the project director or headquarter
staff and the regional head office of evaluation should agree in the selection of supervisors. The
regional office may be an institution such as a science centre or a university which has
administrative support. Close links with the central group and a co-ordinated work environment
are fundamental to the study's success. It can also have decisive consequences for later phases
of the project.

In the integrated science project, several steps were taken to maintain this atmosphere;
central group personnel would visit the regional groups and a general meeting was organized in
which all supervisors participated. These steps were costly; however, results were justified
both in terms of information obtained and group morale and thus they should be included in
evaluation's budget.

Teacher meetings are a source of information that cannot be obtained by means of other
instruments. To create an environment of open discussion is difficult but essential if the meetings
are to fulfil their purpose. The occasional comments provide important data in the reformulation
of the project; therefore they must be noted and added to the other results.

Although the final compilation and analysis of data is not yet completed, the interim results
are available: teacher reports indicate that despite the regional differences, similar student
reactions occur in the varying states; the time spans allocated for each chapter have been
insufficient. There are differences between the average time devoted to each chapter in both the
group with trained teachers and that with untrained teachers; suggestions for changes are very
similar in the three states and it is likely that later projects aiming at formative evaluation will be
carried out in fewer regions, thereby cutting expense while providing an equally accurate evaluation;

the pupils believe they are learning as long as they are memorizing contents and gathering information. Their criterion of teaching and learning remains that of acquiring factual knowledge; the students ignore illustrations which are not referred to by the text; the teachers think that the book determines a change in the methodology of the classes they give. As the pupil text was organized in a way to induce a specific class organization, including team discussions, independent work, general discussion, laboratory and simulations, the teachers follow the instructions and change their roles. Very few teachers resented receiving prepared material which could repress their creativity and freedom to work.

Were the right questions asked? We may suppose that revisions based on the results of the evaluation will make the project more adequate and more interesting to the students. We also learn that the training causes differences in teacher behaviour. However, because it is difficult to control regional training, the analysis should consider the qualitative variations which occur. We may therefore suppose that the courses give greater confidence to the teachers working with the project.

The role of the teachers' guide, an important part in the project, could have been better observed. Nevertheless, its effects will be compared to the results achieved by non-trained teachers.

Has the project affected teacher/pupil relations? Although it would be useful to observe and analyse these, personnel and financial limitations do not permit it. Direct information from teacher reports suggests that there are changes in teacher behaviour.

How are the instruments working? In the pre-test stage, some test questions presented problems. There had been no time to analyse and hence evaluate the questions before testing. As a result, some questions will be eliminated from the analysis leading to the final evaluation. Although initially it was not planned to use tests to grade students, they were utilized for several reasons, including the fact that students are motivated by grades.

As the questions on the questionnaires suppose free answers, their analysis represents a tremendous task. Though the analysis of some information may be useful for review, the coding task is expensive and hard to do.

The attitude scale will be useful for the evaluation, and it may be utilized later in other evaluation studies.

In some cases, the opinions of teachers in their reports did not agree with the author's opinions. The developers wanted to change some topics considered to be traditional although teachers liked them. It was decided to teach traditional subject matter, in a restructured form.

Interim appraisal

The project has already shown results. It has created a more innovative environment. It has brought science teachers and evaluators together and allowed a greater flow of ideas between both groups. Educational researchers have become less abstract and more practical while science teachers have become involved in educational research. It has identified a series of resistance points that must be taken into account during the diffusion phase of the project. It has produced several recommendations concerning the preparation of curriculum projects which may be incorporated by other teams. It has assisted in specifying conditions under which the project is or is not recommended during the diffusion phase, taking into consideration several regional differences that may be useful to regional groups. It has supplied elements to review the project in terms of results, opinions and reactions of students and teachers. It has presented the project to varied audiences serving as the first step of the diffusion phase. It has transmitted to teachers and to administrators the importance of the evaluation for a curriculum development project, in the sense that the material must be subject to classroom testing and analysis on the basis of the results obtained, before being made commercially available.

BIBLIOGRAPHY

Tolman, R., Robinson, J.T., Krasilchik, M. - An evaluation of biology, vol. 1, BSCS 1973 (obtainable from BSCS - P.O. Box 930, Boulder, Colorado).

15 Evaluation of integrated science teaching in Malaysia

Sim Wong Kooi
University of Malaya, Malaysia

SUMMARY

The evaluator has problems enough when he is attempting to assess a course in science which is trying at one and the same time to up-date its content, to introduce an integrated approach, and to shift the emphasis from didactic teaching to learning through active discovery. But his problems are compounded when national policy is seeking to generate an awareness of the importance of science to society and is requiring change in the language of instruction from English to the National Language. This chapter illustrates the problems of evaluation against a complex background of pedagogical reform and of changes in society.

Prologue: a Malaysian scenario

It would not be incorrect to say that Malaysia has been pursuing two virtually antithetical trends. On the one hand, it is trying to keep pace with more advanced countries by embarking upon science and mathematics curriculum reforms on several fronts simultaneously. One avowed objective has therefore been the "improvement of the quality of education for the building of a progressive society oriented towards modern science and technology".[1] On the other hand, in keeping with the trend in newly-independent countries, which strive to establish national identity, Malaysia has also invested in the progressive implementation of its national language (Bahasa Malaysia) policy. By 1975, the medium of instruction for all subjects in the former English-medium primary schools had been converted to Bahasa Malaysia. This has posed problems because the bulk of the qualified science and mathematics teachers are non-Malays who lack either basic proficiency in the language or confidence in teaching in Bahasa Malaysia, or both, and the national language itself is undergoing transition through the accelerated (albeit rather confused) process of coining scientific terminologies. In respect of the many science and mathematics curriculum reforms that were concurrently being envisaged, the writer had urged as early as 1968 that "we should not kid ourselves any longer that evaluation could wait. It is part and parcel of the whole process of curriculum development and it ought to be thought of right from the start".[2]

While the writer is hopeful that the authorities concerned - now that virtually all schools have adopted the new curriculum - would embark upon a full-scale summative evaluation of integrated science which might even constitute a kind of reflective evaluation in order to consider yet further science curriculum reforms, some initial efforts might be illuminating. In subsequent sections,

[1] Second Malaysia Plan, 1971-1975, Kuala Lumpur, Government Press, 1971, p. 232.

[2] Sim Wong Kooi. In a Forum on: "Processes of Curriculum Innovation in Science Education", (31st July 1968), Masaalah Pendidikan, 1969, vol. 2, no. 1, p. 298-311.

apart from indicating desirable directions for the comprehensive, continuous and co-ordinated evaluation of integrated science, a few evaluation projects which have included some aspects of integrated science will be reviewed, even though they fall short of the total expectation.

Overview of implementation efforts

The integrated science curriculum in Malaysia has its origins in Scotland, where the syllabus was meant for a two-year course. The syllabus was modified to meet local needs and additional materials were incorporated for the three years of lower secondary school. Beyond the similarity in name and in the adopted curriculum materials, there is very little resemblance in the context and manner of implementation. In the first place, a prevailing preoccupation with examinations seems to permeate pedagogical practices in Malaysian schools. Results favour teachers who provide copious notes and drill their pupils on past examination questions. Attempts have been made to incorporate questions which go beyond "knowledge" or which relate to practical activities. However, not only are these types of questions no longer drill-proof, but, with the time and personnel constraints in setting examination questions based on a rigid multiple choice format, the proportion of such questions is by no means high. Scottish teachers, on the other hand, are not required to prepare their pupils for a public examination at the end of the two-year integrated science course. Hence, they can provide a less detailed treatment of (or even occasional departures from) the topics (so as not to detract from the basic concepts or principles) than their Malaysian counterparts, who prefer to err on the side of including more details just in case the examination questions stray outside what might normally be expected. Consequently, even when the two-year course is stretched for three years, some teachers have expressed difficulty in "covering" the syllabus.

The lack of incentive or inclination to depart from prescribed procedures is also symptomatic of a system, in which teachers are normally expected to follow directives passively rather than take initiatives for which they may be taken to task. It is doubtful that this attitude of unquestioned acceptance of authority, coupled with the convergent thinking encouraged by the objective-type examinations, would be conducive to the development of inquiry skills in pupils. In contrast, Scottish teachers have had a long tradition of exploring alternatives and deciding on what curricula they wish to adopt. Thus, unlike Malaysia, where all schools will eventually be required to adopt the integrated science syllabus and "trial" merely connotes implementation for minor adjustments without the option of eschewing the entire curriculum if found to be seriously lacking in a particular situation, not all Scottish schools have accepted the integrated science syllabus; of those that have, considerable leeway exists in respect of diversity in implementation.

Another characteristic is that most Malaysian classes exceed forty pupils. On top of this, in most Malaysian schools, the laboratories are relatively poorly-equipped, and typically, teachers of lower secondary classes do not enjoy the luxury of securing help from trained laboratory assistants. With the rather heavy teaching load, only the few more dedicated teachers would be prepared to spend endless out-of-school hours preparing for practical activities. The more enterprising teachers have however mobilized "lab-squads" from among more enthusiastic pupils to assist them not only in preparation and maintenance functions but also in improvisation efforts.[1] Since 1973, the Ministry of Education has doubled the per capita science grant for lower secondary classes and made provision for all schools to have laboratory attendants for these classes. As part of the programme for the implementation of integrated science, many laboratory attendants have attended in-service courses to improve their skills in laboratory maintenance and repair. While no evaluation has yet been made of the efficacy of these courses, many of those attending have indicated that frequently the techniques they learn are difficult to apply because of the lack of suitable equipment in schools or the lack of opportunity to practise the skills. Hopefully, with the programme of capital allocation for basic equipment, the schools will have less cause to

[1] S.K. Pang, "Integrated Science in a Malaysian Context", Sains dan Hisab (a Journal of the Selangor Science Education Association), 1975, p. 20-26.

complain. While the situation may be said to be improving, it is still a far cry from that of Scotland.

Having pointed out some of the salient differences between the context of integrated science in Malaysian schools and its progenitor, it might be useful to examine more closely its nature and mode of implementation. First and foremost, it is important to examine what is meant by "integrated". Whatever the approach, the critical emphasis is on the fundamental unity of science which cuts across traditional subject boundaries. Integrated science in Malaysia, which might be regarded as a mixture of conceptual and environmental approaches, places emphasis on three broad aspects, namely, "developing an understanding of environmental phenomena", "developing a functional understanding of the inquiry skills and modes of thought of science", and "developing a healthy attitude toward the application of scientific knowledge".[1] Scrutiny of the sixteen syllabus sections, however, does not indicate a degree of thematic integration much beyond that which obtained with the former general science syllabus, as shown by the distribution of the topics in Table 26.

Table 26. Distribution of integrated science topics by subject emphases.

Basically 1 subject			Predominantly 1 subject			Equal emphasis on 2/3 subjects	
Biology	Chemistry	Physics	Biology	Chemistry	Physics	Biology and Physics	Biology, Chemistry and Physics
2. Looking at living things 6. Cells and reproduction	9. Hydrogen acids and bases 11. Solutes and Solvents	7. Methods of heat transfer 8. About electricity 13. More about electricity 14. Fluid pressure	12. Transport systems	4. Particles of matter 5. Some common cases 16. The earth	3. Energy 10. Detecting the environment	15. Support and movement	1. Introducing Science

Sections 1-6 are meant for the first year of secondary school (Form I) while sections 7-11 and 12-16 are meant for the second (Form II) and third (Form III) years, respectively. The scheme for progressive implementation is depicted by Figure 15. Although the target date for completion of the implementation to all schools was 1975, a few schools had still not been able to convert to integrated science, mainly because of the lack of trained teachers and equipment and the absence of laboratories. Repercussions on the centralized examination system are likely for it must be very uneconomical to continue to set a separate paper for these few schools.

[1] Kementerian Pelajaran. Integrated Science for Malaysian Schools Form I-III (Provisional Syllabus), May 1975, p. 1.

Figure 15. Progressive implementation scheme for integrated science.

Year:	1969	1970	1971	1972	1973	1974	1975
Form III			22	62	153	174	244
Form II		22	62	153	174	244	232
Form I	22	62	153	174	244	232	199
Total	22	84	237	389	571	650	670

For three successive years, the first batch of integrated science teachers from twenty-two schools attended in-service workshops conducted by a group of Scottish educators. Subsequent groups were trained by local officers who soon found it necessary to decentralize and utilize some from the first batch, together with promising teachers selected on the basis of observations of their performance during visits to schools, as tutors in these in-service courses. The usual practice is for the teachers to go through the worksheets, together with the accompanying activities, and to discuss problems encountered. The duration of these in-service courses is short, being limited to vacations. No systematic evaluation has been carried out concerning the effectiveness of these courses apart from course director reports on their perceptions regarding problems, especially those pertaining to the worksheets and associated activities. Other sources of feedback concerning integrated science teaching have also been used and these are discussed in the next section.

Feedback from formative evaluation

During the formative period of curriculum implementation, some five or six teachers and a few inspectors of schools have been invited regularly to discuss experiences and problems. Based on field observations during visits to schools as well as letters from teachers, especially during the initial stages, slight modifications to the worksheets were made or at least suggested through a newsletter to schools. Unfortunately, the publisher has been reluctant to change the worksheets. It follows too that changes in the activities and the type of equipment and materials used are difficult to effect. Fortunately, though, teachers have been stimulated to continually adapt and improvise through a variety of suggestions made in the Integrated Science Newsletter, which has now been incorporated in Berita Sains, as well as, more recently, through the advent of some rather detailed Panduan Guru (Teacher Guide). The style in the newsletters has been vivid and appealing, with a lively and at times humorous style, and has contributed considerably to boosting teacher morale.

During the initial stages of the implementation of integrated science, enthusiasm was high, not only on the parts of the officers-in-charge but also among the teachers. In contrast, sometime in the middle of 1973, there was a spate of newspaper articles, news items and letters to the press which suggested that all was not well. The report of a Schools Heads' Conference, for instance, stated: "Heads of Secondary Schools today called on the Ministry of Education to suspend the extension of the integrated science, Nuffield Science and the modern mathematics programme until a proper evaluation of these has been made".

A resolution adopted the closing day of the twelfth national conference of nearly 500 heads of Secondary Schools read: "The suspension is particularly important as pupils in SRP/LCE classes

may become innocent victims of the poor pilot scheme and end up as dropouts". [1]

This situation prompted the Ministry of Education to respond by convening a workshop to examine the validity of these statements. Since the bulk of those who attended comprised mainly Ministry officials and teachers who have been recognized for their enthusiasm, it comes as no surprise that after considering a series of point-by-point rebuttals, the inevitable conclusion was reached that the furore was but a storm in the tea cup.

It should have been evident that what was urgently needed would have been a well-planned and well-executed summative evaluation of integrated science. While there are plans for such an evaluation, there has apparently been some reluctance to exploit several possible evaluation avenues, such as careful re-analysis of data collected from periodic feedback questionnaires, scrutiny of several past reports involving the evaluation of some very specific aspects of integrated science and other modern curricula, and capitalizing upon the willingness of some researchers, who have perforce been required to conduct evaluation studies on a reduced scale.

As far as the writer is aware, only one attempt has been made during the initial stages to summarize the potentially useful source of teacher feedback, although more systematic analysis might have provided useful insights for more relevant decisions. As the number of teachers increased in subsequent years, even rather simplistic analysis was not undertaken, partly because it became an increasingly onerous task and the officers-in-charge were not able to derive more meaning from the data other than certain interesting open-ended remarks which were commented upon in the newsletters. Perhaps the assistance of researchers might have been enlisted not only in the data analysis but also in improving the questionnaire format to facilitate processing by computer.

Evaluation vis-a-vis equipment aid

The Ministry of Education is fully aware of the need for schools to be better equipped for teaching practical-oriented courses such as integrated science. Accordingly, in the first set of United Nations Children's Fund (Unicef) aid programmes to the Ministry, science equipment aid formed a major component and the list of equipment sought was geared to the needs of integrated science teaching. Prior to the next phase of Unicef aid to Malaysia, an evaluation of the Unicef aid programmes was undertaken.

Some of the more important findings are presented here. They arise from interviews with officers-in-charge of the supply of science equipment, sending out proformas to all schools and questionnaires to a sample of schools regarding certain aspects of equipment supplied by Unicef and field interviews with principals and teachers who teach or supervise integrated science in schools. [2]

Out of 125 possible items supplied by Unicef, forty were sampled so as to represent three main categories of science equipment, namely, general accessories, such as rubber tubing, thermometers and beakers (thirteen items), specialized accessories, such as platinum electrodes, hydrometers, and microscopes (thirteen items), and specialized units, such as energy conversion kit, human eye model and pH comparator (fourteen items). In order to make more a meaningful interpretation of the data, a number of indices were suggested, showing congruence, in terms of the crude criteria in comparing responses to whether or not the aid was received (RA) with whether or not the school already has the item (AH) or whether or not the item is frequently used (FU), as shown in Table 27.

[1] Straits Times, 9th July 1973.

[2] Sim Wong Kooi et al, *An Evaluation of Unicef Aid Programmes to the Ministry of Education Malaysia, 1968-1972*, ERRU-EPRD, 1971. (A condensed version also appears in *Jernal Bahagian Perancang dan Penyelidikan Pelajaran*, Kementerian Pelajaran Malaysia, 1974, p. 87-139.)

Table 27. Indices of congruence/incongruence.

Type of congruence/ incongruence	(1) Whether received aid (RA)	(2a) Whether school already has aid (AH)	(2b) Whether frequently used
C+	Majority Yes	Majority No	Majority often
C-	Majority No	Majority Yes	Majority not often
I+	Majority Yes	Majority Yes	Majority not often
I-	Majority No	Majority No	Majority often

Alternatively, the indices may also be summarized in the form of the matrix shown in Figure 16.

Figure 16. Indices summarized in the form of a matrix.

		AH (or FU)	
		No (or often)	Yes (or not often)
RA	Yes (or not often)	C+	I+
	No (or often)	I-	C-

When the items were grouped according to the types pf congruence/incongruence as well as the three categories of equipment, it was possible to identify items which displayed the I- type of incongruence and hence requiring urgent attention.

When the responses were compared according to various sub-groups, however, some characteristic patterns became apparent, as shown in Tables 28, 29 and 30.

Table 28. Distribution of congruence/incongruence patterns according to location.

	Comparing RA with AH				Comparing RA with FU			
	I-	I+	C-	C+	I-	I+	C-	C+
Urban	24	2	11	3	13	1	22	4
Rural	13	2	3	22	4	15	12	9

Table 29. Distribution of congruence/incongruence patterns according to sex composition of school.

	Comparing RA with AH				Comparing RA with FU			
	I-	I+	C-	C+	I-	I+	C-	C+
Single sex	23	0	15	2	9	0	29	2
Mixed sex	9	2	3	26	3	16	9	12

Table 30. Distribution of congruence/incongruence patterns according to medium of instruction.

	Comparing RA with AH				Comparing RA with FU			
	I-	I+	C-	C+	I-	I+	C-	C+
National-type (English or Chinese medium)	30	0	6	4	13	1	23	3
National schools (Bahasa Malaysia medium)	7	9	4	20	4	16	7	13

Thus, "in terms of incongruence, the urban, single sex and national-type schools tend to have proportionately more incongruence of the I- type than of the I+ type. Similarly, proportionately more of these types of schools have congruence patterns of the C- type than of the C+ type. It may therefore be concluded that in respect of these schools the majority do not seem to be receiving aid"[1], even when they do not already have equipment or are using the little they have to the fullest. On the other hand, for rural, mixed and national schools, aid has been received by the majority even though many already have sufficient equipment or they do not use the little they have frequently.

While the questionnaires (which were concerned with orders of priority of items and how they were being used) coupled with school visits largely confirm the findings from the analysis of proforma returns, some noteworthy aberrations were observed, such as the following -

"... electrically operated items were sent to schools in certain rural areas that at present do not have electrical supply, although many of the principals expressed the hope that they expect to have electricity in one or two years' time."

"In one state at least, parts of one set of equipment were distributed to different schools who therefore could not utilise the equipment until and unless they obtained the other parts."[2]

Many other aspects, which were related directly or indirectly with equipment aid as well as integrated science, were also considered in the report, but they are quite beyond the scope of the present paper. It suffices however to emphasize that from the point of view of schools, the availability of adequate and relevant equipment contributes a great deal to the likelihood of successfully implementing the practical-oriented integrated science curriculum.

Evaluation vis-a-vis the Second Malaysia Plan (SMP)

From the standpoint of national planning, an over-riding objective concerns the "improvement of the quality of education for the building of a progressive society oriented towards modern science and technology".[3] The Ministry of Education was accordingly interested in its mid-term review of the Second Malaysia Plan (SMP) to ascertain the extent to which the various curriculum reforms in science and mathematics education are contributing towards the attainment of the SMP goals. A team from the University of Malaya was therefore commissioned to carry out such an evaluation. Although it is not possible to discuss all the findings, a few interesting results pertaining to integrated science will be highlighted. The over-all finding was that a major problem confronting the science and mathematics curriculum reform lies in a kind of paternalism which comprises an "attitude that has a tendency to stifle initiative and reinforce over-dependence". The over-all recommendation was also characterized by "Janus-face", which was interpreted as the "maintenance

[1] Sim Wong Kooi et al, An Evaluation ..., op. cit., p. 103.

[2] Sim Wong Kooi et al, An Evaluation ..., op. cit., p. 115-9.

[3] Second Malaysia Plan, op. cit., p. 232.

of an orchestrated perspective, involving having a look in at least two directions simultaneously".[1] These conclusions were discussed in terms of roles in curriculum reform, teacher supply and demand, problems of Bumiputra pupils, curriculum co-ordination and improvement, articulation of objectives, research and evaluation; they were summed up in the form of two frontispiece illustrations showing "Paternalism" and "Janus-face".

The strategy entailed a series of brainstorming session with various groups of Ministry officials, namely inspectors of schools, officers in charge of planning and research and officers involved in public examinations, administering questionnaires to in-service participants and field interviews. The methodology employed was essentially one of comparing the various newly-introduced curricula in science and mathematics and the main concern has been in terms of SMP goals. The present paper reports only a few of the findings pertaining specifically to integrated science.

Among other things, participants at the various in-service courses were asked to name three main objectives of the science/mathematics programmes. The responses of the sixty-five integrated science participants are summarized in Table 31; objectives given by fewer than five respondents have been omitted.

Table 31. Perceived objectives of the integrated science programme.

Objectives	Number of respondents	Per cent
Enhancing interests/motivation	28	43.1
Improving/changing teaching methods	23	35.4
Learning to handle/perform practical activities	18	27.7
Curriculum renewal for national progress	8	12.3
Improving confidence in teaching	7	10.8
Improving knowledge of science	6	9.2
Improving teacher's knowledge of science	5	7.7
Learning by discovery	5	7.7
Curriculum renewal for requirement of business and industry	5	7.7
Improving competence in using aids/apparatus	5	7.7
Integrating various sciences	2	3.1

The respondents were concerned first with affective, then with psychomotor objectives. Interestingly enough, integration was mentioned by only two respondents. When integrated science was considered together with all the other modern science and mathematics curricula, a composite picture of the prevailing conception of the main objectives of all the new programmes could be expressed as follows -

"The new programmes are basically concerned with improving the methods of teaching (such as through in-service training) aimed principally at motivating pupils with new curriculum materials that are in keeping with national progress."[2]

[1] Sim Wong Kooi et al, A Position Paper on the Evaluation of Science and Mathematics Education in Relation to the Mid-term Review of the Second Malaysia Plan (1971-1975), FEUM, 1973.

[2] Sim Wong Kooi et el, A Position Paper ..., op. cit., p. 38.

The teachers were also asked to identify factors which helped or hindered the implementation of the new programmes. A characteristic feature, as is also the case with the other new curricula, is the existence of corresponding positive and negative factors, suggesting that some schools are comparatively well off while other schools are lacking in respect of particular pre-requisites for successful implèmentation. A particularly pronounced bipolar response concerns the availability (41.5 per cent of responses), or lack (53.8 per cent), of apparatus/aids/facilities. An important negative factor is associated with language difficulties (32.3 per cent). Since pupils have difficulty expressing themselves in filling in the worksheets and the examination requires them merely to select from multiple choice responses, the practice of dictating responses, with or without having carried out the practical activities, is unfortunately reinforced. Another important negative factor seems to be the heavy teaching load (26.2 per cent). A directive from the Ministry to schools suggested that, as far as possible, integrated science teachers should not be given more than twenty-five teaching periods per week. However, this did not take into account earlier directives which suggested that teachers with specialized functions (such as guidance teachers, library teachers and physical education teachers) also should not be given more than twenty-five teaching periods per week. Again, this has tended to provide the less diligent and dedicated teachers with some justification for cutting down on pupils' practical activities, especially since they are time-consuming in terms of preparation and instructional time and may even be negatively correlated to performance in the convergent, objective-type paper-and-pencil examination.

Given what seems to be a rather discouraging environment for practising inquiry, we are curious to know what conception teachers had of "discovery". During the field visits, the sample of teachers interviewed were, inter alia, asked to define the term "discovery". The various responses are summarized in Table 32 by classifying them under certain broad headings. The classifications conform roughly to Wittrock's suggested trichotomy in terms of independent, dependent and intervening variables.[1] The independent variables refer to modes of instruction, and the "discovery" method is used as a contrast to such pejorative terms as "traditional", "expository", "teacher-centred" and "didactic". The dependent variable refer to the goals of instruction and the achievement of a particular discovery would be one of these goals. The intervening variables refer to the internal processes of learners involved in an "act of discovery". Processes include integrating the material learned into one's cognitive structure or acquiring a positive concept of self as an autonomous problem solver. The teachers interviewed seem to have touched on all three aspects, although with less emphasis on discovery in terms of dependent and intervening variables.

Further probes however revealed that the prevailing conceptions tended to be rather superficial in nature. For example, in the conception of discovery as an activity, the concern has not been so much in terms of the mental activity involved but often in terms of the associated physical activity. When discussing the practicability of conducting lessons which reflect the discovery approach, most teachers tended to confine their attention to such questions as: How to keep every pupil busy with doing something, given the limited resources? How to complete the syllabus within the limited time available? How to get pupils to work on their own without raising the noise level in the classroom?

Important though these questions are, they do not come to grips with the crucial issues, such as those pertaining to the careful differentiation among different conceptions of discovery and the delineation of specific arrangements needed to ensure that every pupil, as well as the teacher, would be able to engage in meaningful inquiry.[2]

[1] M.C. Wittrock, "The Learning by Discovery Hypothesis", in L.S. Shulman and E.R. Keisler (eds), Learning by Discovery: A Critical Appraisal, Chicago, Rand McNally, 1966, p. 33-75.

[2] Sim Wong Kooi et al, A Position Paper ..., op. cit., p. 80.

Table 32. Conceptions of "Discovery" among sample of teachers interviewed.

	Frequency	(As a percentage)
Discovery defined as:		
(a) Activity		
1. Self activity	44	(28.0)
2. Experimental activity	36	(22.9)
3. Guided activity	26	(16.5)
(b) Learning		
1. Learning by doing	16	(10.2)
2. Incidental learning	4	(2.5)
3. Learning to learn	2	(1.3)
(c) Internal inquiry		
1. Reasoning/relating	19	(12.5)
2. Environmental awareness	7	(4.5)
3. Re-discovery	3	(1.9)

Source: Sim Wong Kooi et al, A Position Paper ..., op. cit., p. 80.

Some current efforts

A number of research studies have recently been launched or have been contemplated by higher degree students, and these promise to provide invaluable evaluation insights into integrated science teaching. A brief description is given of one such study which attempts to examine correlates of student attitudes towards integrated science[1], the data collection for which was completed towards the end of 1976. An overlapping cross-sectional longitudinal design has been adopted, as shown in Figure 17, so as to enable us to investigate possible changes in attitudes as well.

Figure 17. Overlapping cross-sectional longitudinal design.

1975	1976
Form I	
Form II	Form II
Form III	Form III

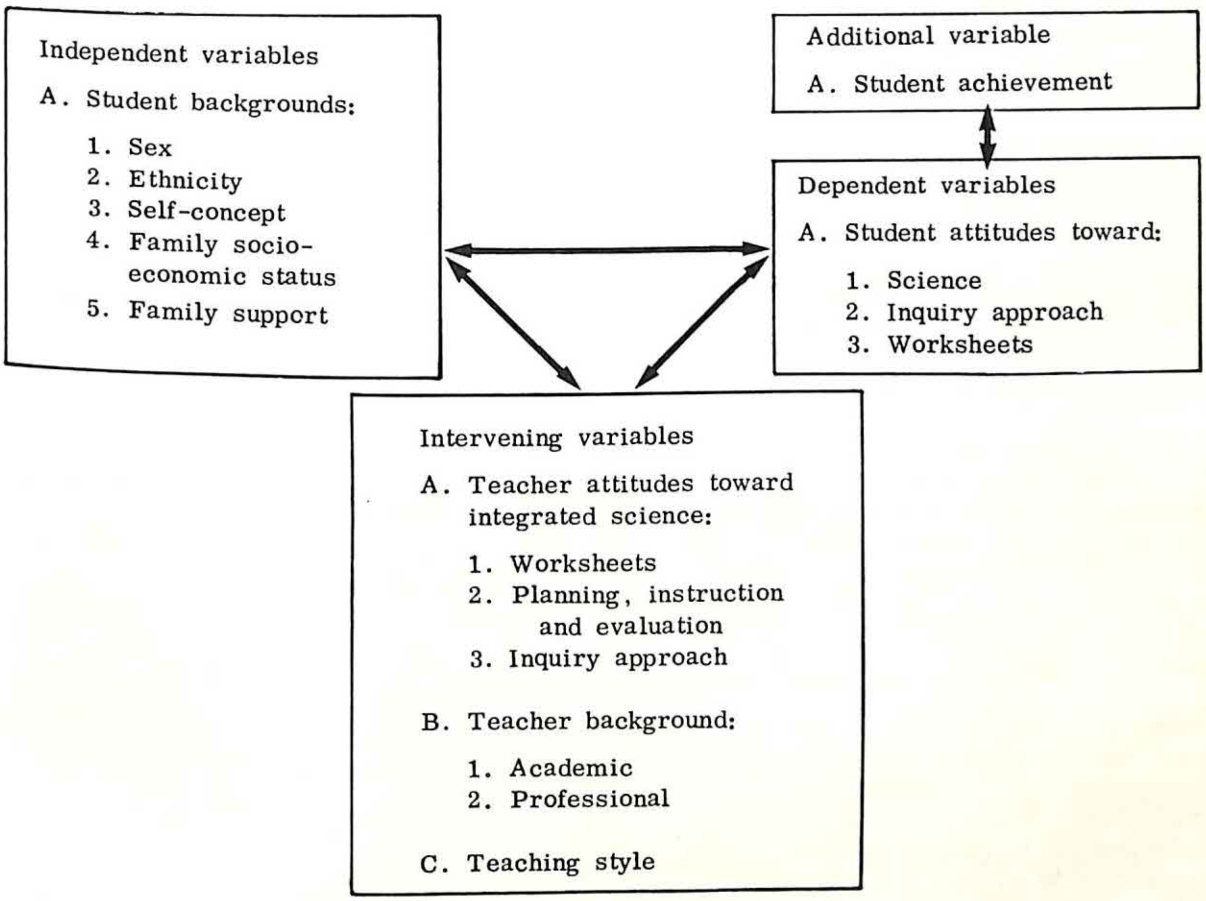
This permits comparisons of the same group of students in different forms as well as different groups of students in the same form. The study attempts to investigate: the attitudes of students towards science in general and towards various aspects of integrated science, in particular the inquiry approach and worksheets; attitude changes of students towards science in general and various particular aspects of integrated science; the relationships between student backgrounds

[1] Lily Suk Men Chew, Private Communication, 1975.

(in particular their sex, ethnicity, self-concept, family Socio-economic status and family support) and their attitudes and attitude changes; the relationships between teacher attitudes, especially towards integrated science worksheets, planning, instruction and evaluation, and the inquiry approach, their academic and professional backgrounds and their teaching styles, and their students' attitudes and attitude changes; the relationships between student attitudes and attitude changes and academic achievement; intercorrelations among student attitudes towards science, inquiry approach and worksheets; intercorrelations among student self-concept, family socio-economic status and family support; and intercorrelations among teacher attitudes, their backgrounds and teaching styles.

The design of the study may therefore be depicted by the diagram shown in Figure 18.

Figure 18. Study design.



Measurement of these variables will be made through:

1. Administration of a student questionnaire, which comprises the following sections:

   A. Background information.

   B. Family support regarding studies, especially in respect of science.

   C. Self-concept regarding studies, especially in respect of science.

   D. Perceptions regarding frequency of various pedagogical practices associated positively or negatively with the inquiry approach.

   E. Attitudes towards science (science teachers and scientists).

2. Administration of a teacher questionnaire, which comprises the following sections:

    A. Teacher background.

    B. Attitudes towards planning, instruction and evaluation of integrated science lessons.

    C. Attitudes towards worksheets, including experiences with each of the sixteen sections.

    D. Attitudes towards inquiry approach.

3. Observations of classroom teaching to determine teaching styles of integrated science teachers. It was hoped initially to observe a sample of classes at different times in 1975 in order to develop a more systematic observation instrument which might provide contrast situations between the type of inquiry[1] approaches adopted by the more indirect and student-centred teachers and those who employ a more direct and teacher-centred approach. This instrument would then be used at various times in 1976, especially at times which are proximate or distal to school or public examinations.

4. Semi-structured interviews of teachers and students. While items would be similar to some of those included in the attitude items in 1. and 2., the interviews would also concentrate on specific problems encountered and general interests and motivations.

5. Student achievement tests, including school and public examinations. If possible, sub-scores, such as for those items corresponding to various categories of Bloom's taxonomy[2] or other alternative classifications associated with or with relevance to practical work and everyday applications, would also be obtained.

With the constraints of time, finance and personnel available, only a sample of English-medium schools in four of the eleven states of Peninsular Malaysia would be included, and these would be further stratified on the basis of year of implementation (1969, 1970, 1971, 1972 and 1973), sex composition (boys, girls and mixed), school size (more than 1,500, 1,000-1,500, and fewer than 1,000) and location (rural and urban). Likewise, classroom observations and interviews with teachers and students would be confined to schools in Selangor.

Epilogue: towards an integrated evaluation strategy

An over-all finding which seems to be emerging from the various evaluation efforts, whether formal or informal, is that the implementation of integrated science has not been integrated into the relevant educational as well as social, cultural, political and economic system in which it has been embedded. Thus, apart from the need to integrate the curriculum horizontally with other curricula which entail approaches diametrically opposed to the so-called inquiry approach, and vertically with developments in science curricula at the primary, upper secondary or even the tertiary levels, it is imperative to consider seriously social expectations based on examination performance, cultural disincentives to question authority, irrevocability of political decisions and the severe limitations of recurrent funds in schools for proper implementation of integrated science.

Likewise, if evaluation is to be meaningful and effective, it should take cognisance of the extent to which intended, as well as unintended, objectives have been attained, and to mobilize all

[1] It is envisaged that, of the possible types of discovery lessons (such as the "Open inductive", "Structured inductive", "Semi deductive", "Simple deductive", "hypothetico deductive" and "transductive" (as explicated by Harold Morine and Greta Morine, Discovery: A Challenge to Teachers, N.J., Prentice-Hall, 1973), only certain types would predominate in our schools while others are likely to be rare.

[2] Bloom et al, Taxonomy of ..., op. cit.

potential manpower and media for the continuous, comprehensive, concerted and co-ordinated evaluation of integrated science. Alternative assessment instruments need to be explored for evaluating important affective, higher cognitive and psychomotor objectives of integrated science. A recently developed simulated practical chemistry test (using slide presentations) appears to be equally if not more effective than the conventional practical test which is costly not only in terms of money but also in respect of time, energy and resources. While such a test is important in assessing pupils' ability to make observations and inferences, practical manipulative skills can only be adequately assessed over a period of time by teachers. Teachers are in fact an important source of manpower for evaluation studies, for they could be encouraged to carry out mini-studies based on detailed specifications of alternative projects. With the increasing interest among teachers in associations for science and mathematics education in Malaysia, it is conceivable that such associations can play a crucial catalytic role. The Curriculum Development Centre can then play the more important role of co-ordinating the various efforts by teachers and researchers, instead of having to undertake ad hoc evaluation studies. To standardize the conduct of criterion tests, it might be possible to exploit alternative media, such as radio and television, for the simultaneous administration of special tests and questionnaires on a nationwide basis.

Some fundamental value questions need also to be critically examined vis-a-vis the desirable purposes, processes and products of introducing integrated science. For example, as depicted by the accompanying photographs of typical laboratory situations, there seem to be abundant indications of pupils busy doing science, but are they really learning science? While the answer may be arrived at empirically, the philosophical, socio-cultural and even political concomitants appear to be many and varied and should be examined or re-examined in concert.

Whether the suggested integrated strategy of evaluation is feasible or even desirable within our predominantly paternalistic system remains an open-ended question, but one that cannot any longer be avoided.

LEARNING OR DOING INTEGRATED SCIENCE ?

16 Evaluation of integrated science teaching in Japan

Hideo Ohashi
National Institute for Educational Research, Tokyo, Japan

SUMMARY

In this case study, student achievement and performance are evaluated in order to obtain useful data for the improvement of the national "Course of Study", curricula and instructional methods. The ensuing discussion delineates the effects of the mode of evaluation on student attitudes toward learning science and indicates the necessity for a more comprehensive kind of evaluation.

Evaluation of the course of study

The "single-track" 6-3-3 school system was inaugurated in Japan in 1947. Attendance at the six-year elementary school is compulsory from the age of six, as is continuation to the three-year lower secondary school for those who have completed elementary school. A child who has completed lower secondary school may go on to an upper secondary school. There are three types of upper secondary school courses: full-time, part-time and correspondence. The full-time course lasts three years, while both the part-time and the correspondence courses last for four years or more. In terms of contents, the upper secondary school courses may also be classified in two categories: general and specialized. Specialized courses are intended to provide vocational or other specialized education for those students who have chosen a particular vocational area as their future career.

The basic standard for school curricula which includes the objectives and teaching materials for different subjects are outlined in the national "Course of study", prepared by the Ministry of Education for each of the three school levels, on the basis of the recommendation of the Curriculum Council, an advisory body to the Minister.

Integrated science is taught from the first grade in elementary school through ninth grade in lower secondary school. The science course in upper secondary school is divided into four areas: physics, chemistry, biology and earth science.

Survey of the scholastic achievement of students

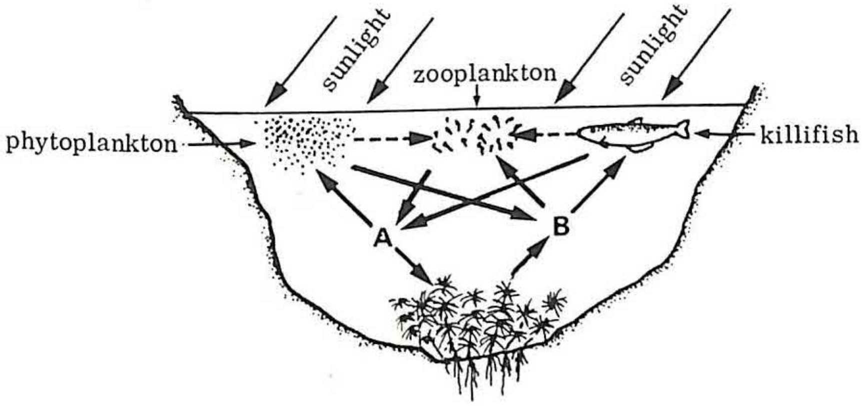
The effects of this new school system and the course of study have been evaluated from different points of view. In particular, heated debates were held among various circles as to whether the scholastic achievement of students is rising or declining. To secure data bearing on this problem, since the early fifties various achievement test programmes have been conducted by prefectural or municipal boards of education, and by public or private educational research institutes.

Since 1956, a nationwide achievement test survey has been conducted annually by the Ministry of Education. The test used for the survey comprises approximately ten questions, totalling thirty sub-questions in all, per subject. The time allotted per subject is fifty minutes. Most of the items are in the form of objective, multiple-choice questions (see example).

Example. (One of the anchor problems in the science test administered to eighth and ninth grades in 1965.)[1]

The following diagram shows how life under water is inter-related. Study the diagram and answer questions 1 and 2 below.

Figure 19. Life under water.



1. Each of the living organisms in the diagram mainly takes in either A or B in the water and breathes out the other during the daytime. The solid arrows show the routes by which they are taken in and out. What substances do A and B stand for?

a. oxygen	b. carbon dioxide	c. nitrogen
d. carbohydrates	e. fat	f. protein

[correct answer: a and b
facility:[2] 26.6 per cent in eighth grade
27.4 per cent in ninth grade]

2. The dotted arrows show the food chain between phytoplankton, zooplankton, and the killifish. Which of the following has the same sort of relationship?

a. root nodule bacteria → pea plant → man

b. worker bee → male bee → queen bee

c. water-flea → killifish → mud snail

d. rice plant → pearl moth → leaf hopper

e. cabbage → grub → sparrow

f. rose → ant → plant-louse

[correct answer: e
facility: 25.5 per cent]

[1] Ministry of Education, Zenkoku Chigakko Gakuryoku-Chosa Hokokusho, 1965, Report on the National Survey of Scholastic Achievement of Lower Secondary School Students in 1965, Research Section, Ministry of Education, Tokyo, 1967.

[2] percentage of students who answered correctly.

The results of the test and their detailed analysis were published every year as an annual report, and distributed to schools and boards of education. Listed below are some important findings accumulated over the past ten years as a result of these surveys.

First, more than one thousand problems with their respective "facility" percentage on the elementary and lower secondary school science have been accumulated. These results are invaluable as a basis for designing science curricula and planning instructional methods.

Next, there are wide variations in the achievement levels of schools. In the case of ninth grade, results showed a discrepancy of eighty points for a 100 point test between the highest ranking school and the lowest ranking school. The standard deviation of the school scores was shown to be around eight points.

In addition, student achievement not only reflects individual abilities and efforts but it is also largely dependent on family background, organization of teachers at school, the size of the school, the availability of facilities, the curricula, urbanization, economic condition and the educational concerns of the community.

There is also a wide range of difference in scholastic achievement, even between schools whose pupils are from similar educational backgrounds and economic and social conditions.

Finally, other points include the fact that a student's achievement on a test is not only determined by the difficulty of the questions but is also dependent largely on the instructional methods.

Although this nationwide survey was proven to be an invaluable source of information, much criticism has been voiced concerning its administration. In spite of the fact that this investigation was conducted on the basis of a paper-and-pencil test so that only a limited profile of the students can be obtained, there is a tendency among educators to attach an unduly high significance to the results. In particular, these surveys and evaluations at schools tend to be summative in nature. Many educators and researchers feel that more efforts should be made to incorporate "formative evaluation" as well. If in fact such surveys are conducted continually year after year, there is a danger that too much emphasis will be placed only on acquiring knowledge. Also, it might create unecessarily tense classroom situations. The information obtained by sampling a large rather than limited number of schools does not merit the additional expense. For all these reasons, large-scale surveys were suspended after 1967. Since 1975 surveys have been resumed by the National Institute for Educational Research, but on a smaller scale at a number of selected schools and with diagnostic test problems on some specific topics.

Survey of teachers' opinions

In revising the course of study, considerable importance is given to teachers' opinions. The Teachers Associations formalize the opinions of teachers and parents and present the results to the public. The teacher consultants compile reports of the problems in education and the opinions of teachers which are reported to the boards of education for discussion. The prefectural and municipal educational institutes also contribute in conducting investigations of teachers' views. These institutes engage in the in-service education of teachers and also conduct practical research in close co-operation with teachers. Their staff members are united in forming a National Federation of Educational Research Institutes in Japan.

The Federation conducted a nationwide survey of educators' opinions in 1970[1] to obtain basic information necessary for the improvement of compulsory education. It wanted to determine the aspects of the current educational system which were considered problematic by those concerned and to establish the sort of revisions which were considered necessary. In terms of compulsory education, considerable emphasis and interest were focused on the best means of developing each individual child's abilities.

[1] National Federation of Educational Research Institutes in Japan, Gimukyoiku Kaizen ni kansuru Iken Chosa. (Report on the Survey of the Opinions concerning the Improvement of Compulsory Education), National Institute for Educational Research, Tokyo, 1971.

Some fifty questions were drafted dealing with such basic factors for the development of children's abilities as the problems of: child development, educational systems including grade systems and class formation, the scope and sequential ordering of instructional materials and the diversity of teaching methods and pre-service and in-service teacher training.

The following is one of the fifty questions:

The course of study contains outlines of materials to be taught. As a principle, every pupil is expected to master the material. Under this system, what percentage of the pupils do you think have an over-all understanding of the material?

1. Over three-quarters of the pupils.
2. About half of the pupils.
3. About one-third of the pupils.
4. Less than one-quarter of the pupils.
5. Uncertain.

The questionnaire was distributed to teachers in 727 randomly-selected elementary schools and in kindergartens and lower secondary schools in the same school districts as the elementary schools, to teacher consultants on boards of education and to researchers at research institutes. Table 33 shows the distribution of the responses.

Table 33. The degree of comprehension of the material. (Expressed as a percentage.)

	No response	1. Over three-quarters of the pupils	2. About one-half of the pupils	3. About one-third of the pupils	4. Less than one-quarter of the pupils	5. Uncertain	Total (actual number of responses)
Elementary school teachers	1.1	28.9	49.2	14.0	2.2	4.5	1,591
Lower secondary school teachers	1.1	16.7	50.2	26.1	4.1	1.8	1,884
Teacher consultants	1.4	29.9	50.5	10.8	0.7	6.6	2,361
Staff members of prefectural educational institutes	1.4	20.1	50.0	12.8	1.5	14.2	1,032
Staff members of municipal educational institutes	2.0	20.6	43.8	20.9	3.3	9.5	306

Generally, speaking, the greatest number of responses was that about half of the pupils understand the materials. Nevertheless, a considerable number of educators responded that less than one-third or one-quarter of the pupils understand the materials. Table 34 shows the results obtained on the degree of comprehension of lower secondary school pupils for each different subject.

Table 34. The degree of comprehension of the material (for courses taught in lower secondary schools). (Expressed as a percentage.)

	No response	1. Over three-quarters of the pupils	2. About one-half of the pupils	3. About one-third of the pupils	4. Less than one-quarter of the pupils	5. Uncertain
No response	10.2	16.3	45.9	23.5	3.1	1.0
Japanese Language, Social Studies	0.2	16.4	51.2	27.3	3.4	1.7
Mathematics, Science	-	15.4	53.8	26.2	3.6	0.8
Music, Fine Arts	2.2	21.3	40.4	25.0	6.6	4.4
Physical Education, Health	-	17.8	45.0	30.2	4.7	2.3
Home Economics, Industrial Arts	0.6	21.9	54.8	18.7	1.9	1.9
Foreigh Language	0.5	13.3	51.7	28.9	4.3	0.9
Other subjects	1.7	16.5	47.5	25.0	6.4	3.0

It can be said that lower secondary school teachers have a greater degree of awareness of this problem than do elementary school teachers. There do not appear to be major subject related variations.

The results of the investigation above is an explicit representation of the opinions of educators. In this sense, it may be considered that there are no new findings on this topic. Yet is is indeed a grave problem that such a large number of pupils are unable to follow the lessons.

Integrated science course in upper secondary schools

In the twenty-five years since the adoption of the new educational system, three successive revisions have radically altered the course of study. One of the driving forces of the revisions seems to have been the increase of enrolment in schools. As shown in Table 35, enrolment in upper secondary schools rose from 42.5 per cent in 1950 to over 90 per cent by 1974. It is also interesting to note the great popularity of pre-school education - over 60 per cent of the children go to kindergartens and a considerable number go to day nurseries. Children have many opportunities to learn prior to entering elementary school.

Table 35. Proportion of the age group enrolled in schools. (Expressed as a percentage.)

School level	Age range	in 1950	in 1974
Kindergarten	3-5	8.9	61.9
Compulsory (Elementary and Lower Secondary)	6-14	99.4	99.9
Upper Secondary	15-17	42.5	90.8
Higher education	18-21	10.1[1]	34.7

[1] 1954.

In 1973, a new science course called Basic science was created as one of the subjects for upper secondary school science programmes. The course was designed originally for non-science students, to give them a comprehensive idea about the nature of science, the scientific way of thinking and the relationship between science and culture. Materials selected from physical, chemical, biological and geological sciences are integrated through the concept approach or the topic approach. However, the course has been in minimal demand; only 2 per cent of the upper secondary school students took the basic science in 1975, while 29 per cent took physics, 35 per cent chemistry, 28 per cent biology and 14 per cent earth science. The low demand for the basic science course may be related to the question of re-educating teachers for this new course and to the need for revising external examination procedures.

Evaluation of performance in experimentation

There are many factors contributing to the complexity of practical learning situations. The teachers' approach to these factors seems to be oriented more toward synthesis in comparison to the researchers who tend to deal with them analytically.

Recently, great emphasis has been placed on the importance of the "inquiry approach" and "process skills" in the teaching of science. It is becoming increasingly popular to have the students actually observe natural objects and phenomena and engage in experimentation and observation as an important part of the learning activity.

The following is a summary of the aspects of the traditional tests which teachers feel are problematic. They feel that a paper-and-pencil test can be used to evaluate a student's knowledge, but is insufficient to evaluate ability and attitude. Furthermore, the student responds to the objective tests only by selecting from the given items. The objective tests are convenient where large numbers of answer sheets need to be processed and when results are dealt with statistically. But in the case of only one class, students' reports and essays written under little surveillance are probably more effective for obtaining educationally valuable information.

With these ideas in mind, various kinds of practical research are being carried on. An example of a project in the development of performance evaluation undertaken through the close co-operation of school teachers and staff members of the Prefectural Education Centres follows.

Finding out the difference of the concentrations of salt solutions

First, a problem such as the one shown below is given to the children to answer in written form (25 minutes).

> Beakers A, B and C contain salt solutions of varying concentrations. What methods exist in ordering the beakers according to decreasing degree of concentration? List as many ways as possible to investigate this problem, using a diagram when necessary.

A summary of the answer given by fourth, fifth and sixth graders in an elementary school to this problem is shown in Table 36.

Table 36. Proposed methods of differentiating the concentrations of salt solutions: difference depending on the grades of the children. (Expressed as a percentage.)

Number	The methods proposed by the child	Fourth grades	Fifth grades	Sixth grades
1	Taste the solutions	94	100	99
2	Evaporate the solutions and compare the residues	24	15	74
3	Check the conductivity of electricity of the solutions	-	-	60
4	Test the buoyancy of objects in the solutions	11	63	39
5	Measure the weights of the solutions	14	42	29
6	Compare the weights of the solutions	8	28	12
7	Compare the colours of the solutions	36	63	38
8	Compare the colour reactions of litmus paper	-	-	35
9	Pour the solutions over slugs	26	8	20
10	Add a fixed amount of salt and compare the amounts undissolved	11	13	10
11	See how much more can be dissolved	3	36	8
12	Compare the solutions by feeling the textures	16	7	6
13	Use a hydrometer	-	-	1
14	Other methods	(16)	(18)	-

From Table 36 it is possible to see whether the children considered this problem based on learning experiences in their school or in light of their daily experiences.

Next, the solutions A, B and C are distributed to each child and they are told to carry on the investigation through actual experimentation. The order and choice of methods of the experiments are left up to the individual child. All the observations made during the experiments are to be recorded by the child as the experiment progresses. Approximately forty minutes later, when the children are finished with the experiments and recording, the report sheets are collected, and the experimental results are summarized in a list. Table 37 was thus compiled from the data produced by twenty-one boys in the sixth grade.

Apropos this list, in investigating the relationships between the three items on the right of Table 37, the following results are obtained.

Table 37. Proposals and results of experiments.[1]

The methods proposed by the child / Children	1. Taste the solutions.	2. Evaporate the solutions and compare the residues.	3. Check the electric conductivity.	4. Test the buoyancy by the solutions.	5. Measure the weights of the solutions.	6. Compare the weights of the solutions.	7. Compare the colours of the solutions.	9. Pour the solution over slugs.	10. Add more salt, compare the amount undissolved.	11. See how much more can be dissolved.	12. Compare the solutions by feeling textures.	14. Other methods.	The number of methods proposed by the child.	The number of successful experiments.	Evaluation scores by the paper-and-pencil tests.
1	⊙	o	⊙	o		⊗	o	o			o		8	2	3
2	o	⊙	⊙			⊗	o				o		6	2	3
3	⊙		⊙	⊙		o	△			⊙			5	4	4
4	⊗	o	o	o					◯	◯			4	2	4
5	⊙	⊙	⊗	◯	◯	o	o	o		◯	o		7	5	4
6	⊙	⊙	△o		◯								3	3	2
7	⊗	o	⊗	o	⊗			o					6	0	2
8	o	o	⊙				◯		o		o		5	2	3
9	⊙	⊙	△o	o	o	×		o			o		7	2	5
10	⊙	o	△		×								2	1	1
11	⊙	o	◯	o			o	o				iron rust △	5	3	4
12	△o	△o	△o	o			o						5	0	2
13	⊙	⊙	⊙		◯							flash light bulb ◯	3	5	2
14	⊙	⊙	⊙		o			o					5	3	3
15	⊙	o	⊙		△o		o	o				odour	6	2	3
16	⊙	o	⊙	◯	o	◯		o		◯			5	5	4
17	⊙	⊙	o				o						4	2	1
18	⊙	⊙	⊙		△o				o				5	3	4
19	⊙	⊙	⊙	⊙				o				making pickle ◯	5	5	3
20	⊙	o	⊙	o	o	◯	o	o				freezing mixture	7	3	4
21	⊙	◯	⊙	o	◯	o	o	o		◯			6	5	3

o Methods proposed in the original investigation. ◯ Successful experimentation.
△ Ambiguous result through experimentation. × Unsuccessful experimentation.

[1] Proposals number 8 and 13 are not included in this table.

The situation is quite similar for girls. That is, the number of methods proposed and the scores are related, but there is little correlation between the number of methods proposed and the number of successful experiments, or between the number of successful experiments and the paper-and-pencil test scores. Thus, it can be said that the learning experience the student undergoes through experimentation cannot be valuated by a paper-and-pencil test, and as can be seen in this example, the reports made by the students pertaining to the experiment are an important basis for evaluation.

The methods explained above require considerable time and expertise on the part of the teacher in analysing the students' reports. There are no appropriate ways to quantify objectively the results of the evaluation. However, the students' observations provide certain intriguing information. Listed below are a few examples.

When the tasting method was used most children were able to identify the densest solution C, but they were divided on the ordering of A and B. One child mentioned in his report that "The sense of taste of the tongue is unreliable".

There were some students who only observed residue, while others measured its weight. Those who could not discriminate using this method did not use enough solution.

The methods employing the conductivity of electricity obtained the most accurate results. Though 90 per cent of the boys proposed this method, only 50 per cent of the girls did so.

The impressions concerning these experiments included such opinions as: I think it's great that we can draw up our own plans for the experiments and use whatever kinds of equipment necessary to carry them out (a girl). What I expect will happen does not always come true when I actually do it (a boy).

It is a very rewarding experience since it is possible to find out the students' ideas and points of view during the process of forming a list by reading off each student's reports. The students welcome this method very much, and do not consider themselves being tested. They take avid interest in the project matter and actively work on it (a teacher).

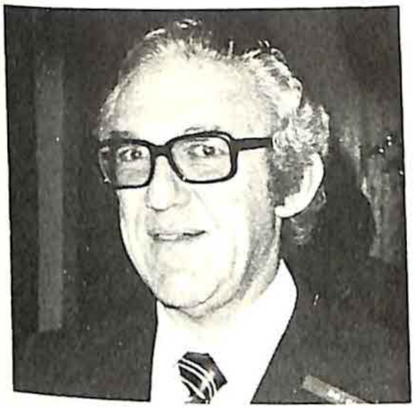
According to the teachers, what they need are examples based on actual experimental results and ways of thinking and methodology suggested by these examples, rather than mere ready-made test instruments. The individual teachers can take these factors into consideration and work out their own system of evaluation. Thinking along these lines, they are devoting their energies particularly to ways of evaluating performance in experiments in the integrated science courses.

BIBLIOGRAPHY

Ministry of Education, National Surveys of Scholastic Achievement, Research Bureau, Ministry of Education, Tokyo, 1961.

M. Ono, K. Sakuma, K. Takeda and A. Minato, Rika Pahomansu Testo. (Performance Test in Science), Toyokan Publishing Co., Tokyo, 1975.

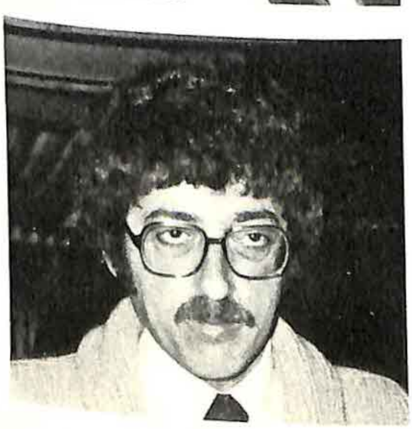
Notes on contributors



DAVID COHEN is Associate Professor of Education at Macquarie University in Sydney, Australia. After teaching science and chemistry for more than 10 years in secondary schools in Victoria, Dr. Cohen became involved first in primary and technical teacher education, and then in curriculum research and development. He has edited a number of journals and published numerous books, papers and monographs concerning science education, curriculum studies, and evaluation. His international activities include consultancies in the Philippines, the United Kingdom with Schools Council, the United States of America (including the Universities of Maryland and Iowa and Michigan State University), France, Poland and Israel. He is a member of the Australian Psychological Society and is currently a Council Member of the Australian Curriculum Development Centre.

DAVID TAWNEY is director of a group at Brunel University developing school science apparatus for the Consortium of Local Education Authorities for the Provision of Science Equipment. After qualifying as a teacher, Mr. Tawney taught science for 13 years in secondary schools in the United Kingdom and Mauritius. Then he helped educate future science teachers at Keele University; while there, he directed an evaluation study of Schools Council Project Technology and was a Topic Leader in the Science Teacher Education Project. More recently, he joined a team funded to evaluate the National Development Programme in Computer Assisted Learning and based at the Centre for Applied Research in Education of the University of East Anglia prior to taking up his present position.

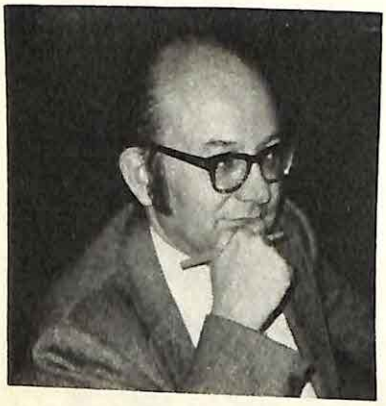
WAYNE W. WELCH is a Professor of Educational Psychology at the University of Minnesota. His major areas of interest include measurement, evaluation and science education. A former physics teacher, Dr. Welch has been involved in the evaluation of a number of science curricula including Harvard Project Physics, Physical Science for Non-Science Students, Ginn Elementary Science Programme, and most recently Unified Science and Mathematics for Elementary Schools (USMES).
He is a Past President of the National Association for Research in Science Teaching and currently serves on the editorial board of The Science Teacher and the Journal of Research in Science Teaching. Under a grant from the National Science Foundation, Dr. Welch is at present completing a study of utilization of new curricula by science and mathematics teachers.



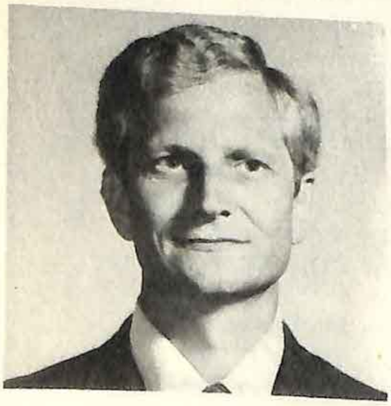
JAN R. BLOCH is a Research Associate at the Institut fur die Padagogik der Naturwissenschaften (Institute for Science Education), University of Kiel, Federal Republic of Germany. After completing his formal education in chemistry, Dr. Bloch was involved in research and teaching for some years until accepting his present position in 1971.

His current interests include the role of communicative competence in science teaching and learning, interdisciplinary science education with an emphasis on the social aspects of both science and technology, and the philosophical and historical bases of science.

CLIVE SUTTON is Senior Lecturer in Education at the University of Leicester. Dr. Sutton worked as a research biochemist before becoming a secondary school teacher in the United Kingdom, and later in India. At various times he has taught single subject chemistry, physics, elementary biology, 'general' science in the primary and early secondary school, and science in the context of general studies for 16-18 year olds. As one of the Co-ordinators of the Science Teacher Education Project, he is senior editor of The Art of the Science Teacher and joint author, with J.T. Haysom, of Innovation in Teacher Education, McGraw-Hill (1974).



VICTOR J. MAYER is Professor of Science Education and Geology at the Ohio State University. He has had a long-time involvement with developments in the teaching of Earth Science, beginning in 1960 with public secondary school teaching, continuing through his doctoral work in Science Education at the University of Colorado, his position as Assistant Professor of Earth Science at the State University College at Oneida, New York, and his current position. Internationally, he has had a special interest in science education in Latin America, having assisted in developing a teacher education programme in Earth Science in Venezuela and in working with science educators from seven other countries of that continent.



JAMES M. RICHMOND is currently a University Fellow at the Ohio State University working on a Ph.D. in Science Education. For his dissertation research, he is conducting an extensive national survey in the United Kingdom to determine the environmental knowledge and attitudes of students at the secondary level. His prior twelve years of teaching experience have been in his native Australia, the United Kingdom and the United States, where he was in charge of the biology programme at Culver Military Academy. At Ohio State University, he has worked as a Research Associate in the ERIC Information Analysis Centre for Science, Mathematics, and Environmental Education, and as a Teaching Associate in the field-based science teacher education programme.

GREGOR A. RAMSEY is Director of the Torrens College of Advanced Education in South Australia. After teaching science and chemistry in secondary schools, Dr. Ramsey spent over two years as Assistant Director of the Australian Science Education Project and has worked with the ERIC Centre for Science Education at Ohio State University (United States). He has published widely in educational journals, especially in the areas of science education and curriculum.

J.F. KERR is Professor in the School of Education at the University of Leicester, where he has responsibility for the further professional education of teachers and is Senior Tutor for the B.Ed. Degree. His publications are in the fields of science education and curriculum development. He is President of the Association for Science Education in the United Kingdom.

WYNNE HARLEN has since 1973 been Director of the "Progress in Learning Science" Project based at Reading University and sponsored by Schools Council. After graduating in physics at Oxford in 1958, Dr. Harlen taught science in schools and as a college of education lecturer. During this time she became particularly interested in science at the primary school level and undertook research into the effect of school experience in science on the development of science concepts in young children. In 1966 she joined the School of Education of Bristol University as research fellow and has been working in the field of evaluation in science curriculum development since that time. From 1967-73 she was engaged in the formative evaluation of the Science 5/13 project, becoming the co-author of the guide to science 5/13 'With Objectives in Mind' and the author of 'Science 5/13: A Formative Evaluation'.

EMMANUEL AYOTUNDE YOLOYE, B.Sc., P.G.C.E., Dip.Ed. (London), M.A., Ph.D. (Columbia, N.Y.) is Professor of Education and Director of the International Centre for Educational Evaluation (ICEE) in the Institute of Education University of Ibadan. After lecturing in Science from 1961, Dr. Yoloye directed the evaluation of the African Primary Science Programme from 1967 to 1971. With the ICEE team he has also carried out evaluation of the Pilot Rural Education Project in Namutamba Uganda, the Primary Science Programme in the Bendel State of Nigeria, the University of Ife, Nigeria Six-Year Primary Project, and the Primary School Mathematics Programme in Mauritius.

JUDITH REAY is Senior Lecturer with the in-service Diploma in Education programme at the University of the West Indies in Trinidad and is closely associated with the science teachers' association movement in the Caribbean. Originally from Fiji, Miss Reay has worked in New Zealand, Fiji, Switzerland, the Caribbean, the United Kingdom, the United Republic of Cameroun and Pakistan. While at the University of the West Indies in Jamaica, she directed the Science Education Project and the Science Centre for Schools, shared in the development of both the MONA and WISC curricula and acted as a WISC consultant for the Northern Caribbean.

A.W. JEFFREY is the senior inspector for science in the Scottish Education Department. After graduating in Chemistry from the University of St. Andrews, he undertook research in industry before entering teaching. He taught for some years in George Watson's Boys College (Edinburgh) and later was Principal Teacher of Science in Madras College, St. Andrews. He joined Her Majesty's Inspectorate of Schools in Scotland in 1963 and has been involved in curriculum development at national level continuously since then. He was seconded to the Malaysian Ministry of Education 1966-67 to advise on science education and has recently undertaken a similar task in Hong Kong.

SALLY BROWN is Senior Research Fellow in the Education Department at Stirling University, Scotland. Dr. Brown is currently engaged on a research project, involving classroom observation of teachers of integrated science, which is concerned with how teachers interpret and practice 'guided-discovery methods', 'integration of subjects', 'differential treatment of pupils in mixed ability classes' and 'teaching towards objectives'.

Her previous experience includes lecturing in physics at the Nigerian College of Technology, the University of Ife in Nigeria and Avery Hill College of Education in London, six years as a science teacher in a Scottish secondary school and research into pupils' attitudes to science in the early years of secondary school.

MYRIAM KRASILCHIK is the Director of the Science Teaching Centre of Sao Paulo - CECISP, and is teaching Science Education and Curriculum Development courses at the School of Education of the University of Sao Paulo. Dr. Krasilchik's past experience include the role of Co-ordinator for the adaptation of the Biological Science Curriculum Study and Nuffield O level-Biology project, and the development of several Brazilian science teaching projects such as IBECC's General Science, "The Scientists" and Integrated Science.

SIM WONG KOOI is Professor of Education at the University of Malaya. He is a graduate from the University of Malaya in Singapore with a B.Sc. (Hons) in Chemistry and a Diploma in Education. After graduating he spent many years teaching in Malaysia following which he was granted a Commonwealth scholarship to study at Melbourne University, Australia. He then went to Harvard University where he took his Ph.D. in Education. He has worked as Unesco consultant in various regional projects on science education and has published many research articles, in particular on evaluation and testing in science education.

HIDEO OHASHI is Director of the Science Education Research Centre at the National Institute for Educational Research, Japan. He began his career as a teacher at Oizumi Normal School and then at Tokyo Women's Higher Normal School. He subsequently moved into the Ministry of Education, rising from Textbook compiler first, to Specialist of science education, then to Deputy Chief of the Secondary Education Section, Ministry of Education, and later to become an Inspector in the Ministry. In 1971, he became Director of the Fifth Department of the National Institute for Educational Research. His special fields of interest are in the training of science teachers, the designing of science equipment and policy on science education.

ISBN 92-3-101440-4